



Marginal quantile regression for varying coefficient models with longitudinal data

Weihua Zhao¹ · Weiping Zhang² · Heng Lian³

Received: 8 October 2017 / Revised: 24 July 2018 / Published online: 18 August 2018
© The Institute of Statistical Mathematics, Tokyo 2018

Abstract

In this paper, we investigate the quantile varying coefficient model for longitudinal data, where the unknown nonparametric functions are approximated by polynomial splines and the estimators are obtained by minimizing the quadratic inference function. The theoretical properties of the resulting estimators are established, and they achieve the optimal convergence rate for the nonparametric functions. Since the objective function is non-smooth, an estimation procedure is proposed that uses induced smoothing and we prove that the smoothed estimator is asymptotically equivalent to the original estimator. Moreover, we propose a variable selection procedure based on the regularization method, which can simultaneously estimate and select important nonparametric components and has the asymptotic oracle property. Extensive simulations and a real data analysis show the usefulness of the proposed method.

Keywords Longitudinal data · Quadratic inference function · Quantile regression · Varying coefficient model

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10463-018-0684-7>) contains supplementary material, which is available to authorized users.

✉ Weihua Zhao
weihuazhao66@hotmail.com

¹ School of Science, Nantong University, 9 Seyuan Rd, Nantong 226000, China

² Department of Statistics and Finance, University of Science and Technology of China, 96 Jinzhai Rd, Hefei 230026, China

³ Department of Mathematics, City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, Hong Kong

1 Introduction

Besides parametric models, various semi-/nonparametric models have been used to describe longitudinal data. The varying coefficient model (VCM), firstly proposed by [Hastie and Tibshirani \(1993\)](#), has been a popular modeling approach for longitudinal data due to its flexibility and interpretability, where the coefficients are smooth nonparametric functions of the measurement time. The varying coefficient model for longitudinal data can be written as

$$y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\alpha}(t_{ij}) + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (1)$$

where $\mathbf{X}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ is a p -dimensional covariate associated with the j -th measurement for the i -th subject at time t_{ij} , $\boldsymbol{\alpha}(t) = (\alpha_1(t), \dots, \alpha_p(t))^T$ comprises p unknown nonparametric functions and ϵ_{ij} is the random error.

There exist many studies on model (1) in the framework of mean regression; see, for example, [Fan and Zhang \(1999, 2000\)](#), [Chiang et al. \(2001\)](#), [Huang et al. \(2002\)](#), [Qu and Li \(2006\)](#), [Şentürk and Müller \(2008\)](#). Quantile regression ([Koenker 2005](#)) is a valuable alternative to least squares-based method. [Kim \(2007\)](#) and [Cai and Xu \(2008\)](#) proposed the quantile VCM for cross-sectional data using polynomial splines and local polynomial smoothing, respectively. [Andriyana et al. \(2014\)](#) investigated the quantile VCM for longitudinal data through the penalized splines approach. Compared with mean regression for VCM, quantile regression can provide a more comprehensive summary of the response distribution and describe the dynamic functional relationship between covariates and response at different percentiles of the distribution. In addition, quantile regression is more robust than least squares regression.

For longitudinal data analysis, it is important to take into account the correlation properly within each subject, and ignoring such correlations could yield less efficient estimators. Some strategies have been developed to incorporate the correlation to fit the longitudinal data in mean regression. [Lin et al. \(2007\)](#) studied the VCM based on generalized estimating equation (GEE, [Liang and Zeger 1986](#)). [Wang et al. \(2005\)](#) and [Huang et al. \(2007\)](#) considered the efficient estimation for partial linear semi-parametric models, and [Lian et al. \(2014\)](#) investigated partial linear additive model in high dimensions. Although GEE-based estimators are consistent, they lose efficiency if the working correlation is misspecified. To alleviate the impact of correlation misspecification and improve estimation efficiency, [Qu and Li \(2006\)](#) applied quadratic inference functions (QIF, [Qu et al. 2000](#)) to VCM using penalized splines. The QIF approach takes into account the within-cluster correlation and is more efficient than the GEE approach when the working correlation is misspecified. Some studies have also been carried out for other nonparametric/semiparametric models using QIF; see, for example, [Xue et al. \(2010\)](#), [Li et al. \(2014\)](#), [Ma et al. \(2014\)](#).

Quantile regression has been widely used in longitudinal data. However, most of these works proposed estimators that ignore the correlation within subject for simplicity, which resulted in low efficiency for estimation and inferences. Compared to mean regression, it is more challenging to account for correlation for quantile regression due to some computational issues as pointed out by [Leng and Zhang \(2014\)](#). In view

of the flexibility of VCM and good performance of QIF, in this paper, we study the estimation and inference problem for quantile VCM. By using the polynomial splines to approximate the nonparametric functions and taking into account the correlation, we propose a new estimation procedure for longitudinal data and establish theoretical properties of the resulting estimators. To further improve estimation, we develop a method to select the important variables using the adaptive group SCAD penalty. The theoretical challenge largely lies in the non-continuity of the objective function as well as the diverging dimensionality resulting from spline approximation, which is dealt with using empirical processes theory (Van der Vaart 2000). Simulation results and real data analysis show that our proposed method outperforms the existing methods.

The rest of the paper is organized as follows. In Sect. 2, we present the estimation approach to quantile VCM based on QIF, where the nonparametric functions are approximated by polynomial splines. The large sample properties of the proposed estimators are established, and we also develop an estimation procedure using induced smoothing. In Sect. 3, to select the important nonparametric components, we develop a variable selection procedure based on the adaptive group SCAD penalty, and its oracle properties are also investigated. Simulation studies in Sect. 4 and real data analysis in Sect. 5 are used to illustrate the performance of the proposed approach. Finally, some concluding remarks are given in Sect. 6. Technical proofs are contained in Appendix in Supplementary Material.

2 Methodology

2.1 Estimation based on QIF

We assume in (1) that the conditional τ -th quantile of error ϵ_{ij} given X_{ij} and t_{ij} is zero, i.e.,

$$Q_\tau(y_{ij}|X_{ij}) = X_{ij}^T \alpha_\tau(t_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i,$$

and that the observations are independent across different subjects. For notation simplicity, we omit the subscript τ in $\alpha_\tau(t)$ in the following.

In our estimation procedure, we approximate the smooth functions $\{\alpha_l(\cdot)\}_{l=1}^p$ by polynomial splines (De Boor 2001; He and Shi 1996). We assume without loss of generality that the covariates t_{ij} are scaled to take value in the interval $[0, 1]$. For each $1 \leq l \leq p$, let

$$\xi_{-l,(d-1)} = \dots = 0 = \xi_{l,0} < \xi_{l,1} < \dots < \xi_{l,K_l} < 1 = \xi_{l,K_l+1} = \dots = \xi_{l,K_l+d}$$

be a partition of $[0, 1]$ into subintervals $[\xi_{l,k}, \xi_{l,k+1}), k = 0, \dots, K_l$ with K_l interior knots. A polynomial spline of order d is a function which is a polynomial of degree $d - 1$ in each subinterval and globally $d - 2$ times continuously differentiable on $[0, 1]$. Denote the d -th order B-spline basis as $\mathbf{B}_l(t) = (B_{l1}(t), \dots, B_{lJ_l}(t))^T, J_l = K_l + d$, and its normalized basis as $\sqrt{J_l} \mathbf{B}_l(t) = (\sqrt{J_l} B_{l1}(t), \dots, \sqrt{J_l} B_{lJ_l}(t))^T$. With an abuse of notation, the normalized basis is still denoted by $\mathbf{B}_l(t)$. Then, each unknown

function $\alpha_l(\cdot)$ can be approximated by a linear combination of the normalized basis such that

$$\alpha_l(t) \approx s_l(t) = \sum_{k=1}^{J_l} B_{lk}(t)\gamma_{lk}, \quad l = 1, \dots, p.$$

Define $\boldsymbol{\gamma}_l = (\gamma_{l1}, \dots, \gamma_{lK_l})^T$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_p^T)^T$, $\mathbf{U}_{ij}^T = \mathbf{X}_{ij}^T \bar{\mathbf{B}}(t_{ij})$,

$$\bar{\mathbf{B}}(t) = \begin{pmatrix} \mathbf{B}_1^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{B}_p^T(t) \end{pmatrix},$$

$\mathbf{U}_i = (\mathbf{U}_{i1}, \dots, \mathbf{U}_{im_i})^T$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$. If we ignore the correlation within subjects, we can obtain the estimator by minimizing the following objective function (Kim 2007):

$$G(\boldsymbol{\gamma}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \rho_\tau \left(y_{ij} - \mathbf{U}_{ij}^T \boldsymbol{\gamma} \right), \tag{2}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function. The estimating equation derived from (2) is

$$\sum_{i=1}^n \mathbf{U}_i^T \psi_\tau(\mathbf{y}_i - \mathbf{U}_i \boldsymbol{\gamma}) = 0,$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^T$, $\psi_\tau(u) = \tau - I(u < 0)$ and $\psi_\tau(\mathbf{y}_i - \mathbf{U}_i \boldsymbol{\gamma}) = (\psi_\tau(y_{i1} - \mathbf{U}_{i1}^T \boldsymbol{\gamma}), \dots, \psi_\tau(y_{im_i} - \mathbf{U}_{im_i}^T \boldsymbol{\gamma}))^T$.

As in Jung (1996) and Leng and Zhang (2014), a more efficient estimating equation takes the form

$$\sum_{i=1}^n \mathbf{U}_i^T \boldsymbol{\Gamma}_i \mathbf{A}_i^{-1} \psi_\tau(\mathbf{y}_i - \mathbf{U}_i \boldsymbol{\gamma}) = 0, \tag{3}$$

where $\boldsymbol{\Gamma}_i = \text{diag}(f_{i1}(0), \dots, f_{im_i}(0))$ with $f_{ij}(\cdot)$ being the conditional pdf of ϵ_{ij} , and \mathbf{A}_i is the working correlation matrix that may depend on some nuisance parameters which, however, may be difficult to estimate. To circumvent this problem, we apply the QIF approach by approximating \mathbf{A}_i^{-1} with a linear combination of basis matrices as

$$\mathbf{A}_i^{-1} = \sum_{k=1}^v a_k \mathbf{M}_{ki},$$

where \mathbf{M}_{ki} 's are known symmetric matrices and a_1, \dots, a_v are unknown constants. As stated in Qu et al. (2000), this is a sufficiently rich family that could accommodate,

or at least approximate, many correlation structures commonly used. For example, if the working correlation has compound symmetric structure with parameter ϖ , then A_i^{-1} can be represented as $a_1 M_{1i} + a_2 M_{2i}$, where $a_1 = -\{(m_i - 2)\varpi + 1\}/k_1$, $a_2 = \varpi/k_1$, $k_1 = (m_i - 1)\varpi^2 - (m_i - 1)\varpi - 1$, M_{1i} is the identity matrix, M_{2i} has 0 on the diagonal and 1 off the diagonal. Similar linear representation of A_i^{-1} also holds for the AR(1) correlation structure with appropriate basis matrices. The advantage of the QIF approach is that it does not need to estimate the linear coefficients a_i 's.

In QIF, we use estimating equations defined as

$$S(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n S_i(\boldsymbol{\gamma}), \tag{4}$$

where

$$S_i(\boldsymbol{\gamma}) = \begin{pmatrix} U_i^T \boldsymbol{\Gamma}_i M_{1i} \psi_\tau(\mathbf{y}_i - U_i \boldsymbol{\gamma}) \\ \vdots \\ U_i^T \boldsymbol{\Gamma}_i M_{vi} \psi_\tau(\mathbf{y}_i - U_i \boldsymbol{\gamma}) \end{pmatrix}. \tag{5}$$

Note that there are more estimation equations than the number of unknown parameters in (4). $\boldsymbol{\gamma}$ can be estimated as

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmin}_{\boldsymbol{\gamma}} Q_n(\boldsymbol{\gamma}), \tag{6}$$

where

$$Q_n(\boldsymbol{\gamma}) = n S^T(\boldsymbol{\gamma}) \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\gamma}) S(\boldsymbol{\gamma}),$$

$$\boldsymbol{\Sigma}_n(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n S_i(\boldsymbol{\gamma}) S_i^T(\boldsymbol{\gamma}).$$

As a result, we define the estimator of the unknown functions as

$$\hat{\alpha}_l(t) = \mathbf{B}_l^T(t) \hat{\boldsymbol{\gamma}}_l, \quad l = 1, \dots, p.$$

Remark 1 Note that the estimating Eq. (4) involves the unknown error density $f_{ij}(0)$. In this paper, we adopt the method of [Hendricks and Koenker \(1992\)](#) and estimate $f_{ij}(0)$ by the difference quotient,

$$\hat{f}_{ij}(0) = 2h_n \left\{ \mathbf{X}_{ij}^T [\check{\alpha}(t_{ij}, \tau + h_n) - \check{\alpha}(t_{ij}, \tau - h_n)] \right\}^{-1},$$

where the estimators $\check{\alpha}(t, \tau)$ can be obtained from (6) by omitting the term $\boldsymbol{\Gamma}_i$ at quantile level τ and h_n is a bandwidth parameter tending to zero as $n \rightarrow \infty$. In our numerical studies, we choose $h_n = 1.57n^{-1/3} (1.5\phi^2\{\Phi^{-1}(\tau)\}/(2\{\Phi^{-1}(\tau)\}^2 + 1))^{2/3}$ following [Hall and Sheather \(1988\)](#), where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution. For the sake of simplicity in the proof, we assume that $f_{ij}(0)$'s

are known in the following in order to derive the asymptotic properties of the estimators. This is similar to the approach adopted in the literature, for example in [Leng and Zhang \(2014\)](#). Unfortunately, it seems very hard to establish the asymptotic property when the density is estimated. On the other hand, we believe a very accurate estimator of the density is not very critical in our context, unlike for classical density estimation problems. For example, even if we replace the density by 1, the estimating equation is still consistent. It required further study to see whether the gap in theory can be closed with more technical advancements.

2.2 Theoretical properties

In this subsection, we study the rate of convergence for $\hat{\alpha}_l(t)$. The following assumptions are needed to derive the asymptotic properties of $\hat{\alpha}_l(t)$.

- (C1) The cluster sizes m_i are uniformly bounded for all $i = 1, \dots, n$.
- (C2) The conditional density f_{ij} of ϵ_{ij} is uniformly bounded and bounded away from zero and has a bounded first derivative in the neighborhood of zero.
- (C3) $\alpha_l(\cdot) \in \mathcal{H}_r$, $l = 1, \dots, p$, for some $r > 1/2$, where \mathcal{H}_r is the collection of all functions on $[0, 1]$ whose s -th order derivative satisfies the Hölder condition of order ϑ with $r = s + \vartheta$. The spline order $d \geq r + 1$.
- (C4) The covariates X_{ij} are bounded in probability for all i and j .
- (C5) The conditional distribution of t given $X = \mathbf{x}$ has a density $f_{t|X}$ which satisfies that $0 < c_1 \leq f_{t|X}(t|\mathbf{x}) \leq c_2 < \infty$ uniformly in \mathbf{x} and t for some positive constants c_1 and c_2 .
- (C6) The knots sequences ξ_{l,K_l} for $l = 1, \dots, p$, are quasi-uniform. That is, there exists $c_3 > 0$, such that $\max_l (\max_k h_{l,k} / \min_k h_{l,k}) \leq c_3$, where $h_{l,k} = \xi_{l,k+1} - \xi_{l,k}$ ($0 \leq k \leq K_l$) are the distances between neighboring knots. Furthermore, the number of interior knots $K_l \asymp n^{1/(2r+1)}$.
- (C7) The eigenvalues for each M_{ki} are bounded away from 0 and ∞ .

These conditions are common in the literature; see, for example, [Kim \(2007\)](#), [Huang et al. \(2002\)](#), [Wang et al. \(2009\)](#) and [Xue et al. \(2010\)](#). Conditions (C1) and (C2) are standard assumptions used in longitudinal studies and quantile regression, respectively. Condition (C3) is a smoothness assumption on the nonparametric functions. The boundedness assumption in (C4) is mainly for convenience of proof. Condition (C5) is commonly used in VCM. Condition (C6) is a mild assumption on the choice of knots. Finally, Condition (C7) is satisfied for commonly used basis matrices.

Theorem 1 *Under Conditions (C1)–(C7), there exists a local minimizer of (6) such that*

$$\|\hat{\alpha}_l - \alpha_l\|_2^2 = O_p\left(n^{-2r/(2r+1)}\right), \quad l = 1, \dots, p,$$

where $\|\cdot\|_2$ is the L_2 norm for functions on $[0, 1]$.

Remark 2 The rate of convergence given here is the same as that in [Kim \(2007\)](#) for independent data ($m_i \equiv 1$) and in [Wang et al. \(2009\)](#) for longitudinal data which,

however, ignore the correlations. In practice, the main advantage of the QIF approach is that it incorporates within-cluster correlation by optimally combining estimating equations without estimating the correlation parameters.

Remark 3 Due to that $Q_n(\boldsymbol{\gamma})$ is not continuous, we are not aware of a satisfactory algorithm to compute its minimizer. We propose a smoothing method to obtain a feasible objective function in the next subsection. However, the theoretical results here provide a benchmark so one can compare the rates of the feasible estimator with the current infeasible one. Similarly, in Sect. 3, infeasible estimator is also studied theoretically when penalization is used for variable selection.

2.3 Induced smoothing

It is difficult to solve the estimating Eq. (6) directly, which is caused by the fact that $S(\boldsymbol{\gamma})$ is not continuous. To overcome this difficulty, we apply the smoothing method which has been used in linear quantile regression by Fu and Wang (2012) and Leng and Zhang (2014).

Let

$$\tilde{S}(\boldsymbol{\gamma}) = E \left[S(\boldsymbol{\gamma} + (nh)^{-1/2} \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}) \right], \tag{7}$$

where $h = n^{-1/(2r+1)}$, the expectation is taken with respect to $\boldsymbol{\delta} \sim N(0, \mathbf{I}_q)$ with $q = \sum_{l=1}^p J_l$, and $\boldsymbol{\Omega}$ is a $q \times q$ positive definite matrix. We note that such choices are certainly not the only possibilities. We just followed the literature on induced smoothing and used the multivariate Gaussian distribution for smoothing. The smoothing in some sense is merely used to make sure that the objective function is differentiable. As long as the disturbance is small enough, we should expect \tilde{Q}_n to be close to Q_n and thus should have reasonable performances.

By some simple calculations, $\tilde{S}_i(\boldsymbol{\gamma})$ can be written as

$$\tilde{S}_i(\boldsymbol{\gamma}) = ES_i(\boldsymbol{\gamma} + (nh)^{-1/2} \boldsymbol{\Omega}^{1/2} \boldsymbol{\delta}) = \begin{pmatrix} \mathbf{U}_i^T \boldsymbol{\Gamma}_i \mathbf{M}_{1i} \left[\Phi \left(\sqrt{nh} \frac{y_i - \mathbf{U}_i \boldsymbol{\gamma}}{r_i} \right) - \mathbf{1}(1 - \tau) \right] \\ \vdots \\ \mathbf{U}_i^T \boldsymbol{\Gamma}_i \mathbf{M}_{vi} \left[\Phi \left(\sqrt{nh} \frac{y_i - \mathbf{U}_i \boldsymbol{\gamma}}{r_i} \right) - \mathbf{1}(1 - \tau) \right] \end{pmatrix}, \tag{8}$$

where $\mathbf{r}_i = (r_{i1}, \dots, r_{im_i})^T$ with $r_{ij} = \sqrt{\mathbf{U}_{ij}^T \boldsymbol{\Omega} \mathbf{U}_{ij}}$, $j = 1, \dots, m_i$, $\mathbf{1}$ being the m_i -dimensional column vector with all elements 1, and $\Phi \left(\sqrt{nh} \frac{y_i - \mathbf{U}_i \boldsymbol{\gamma}}{r_i} \right)$ denotes an m_i -dimensional vector with j -th element being $\Phi \left(\sqrt{nh} \frac{y_{ij} - \mathbf{U}_{ij}^T \boldsymbol{\gamma}}{r_{ij}} \right)$.

The smoothing estimator $\tilde{\boldsymbol{\gamma}}$ can be obtained as

$$\tilde{\boldsymbol{\gamma}} = \operatorname{argmin}_{\boldsymbol{\gamma}} \tilde{Q}_n(\boldsymbol{\gamma}), \tag{9}$$

where

$$\tilde{Q}_n(\boldsymbol{\gamma}) = n\tilde{S}^T(\boldsymbol{\gamma})\tilde{\Sigma}_n^{-1}(\boldsymbol{\gamma})\tilde{S}^T(\boldsymbol{\gamma})$$

with $\tilde{\Sigma}_n(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n \tilde{S}_i(\boldsymbol{\gamma})\tilde{S}_i^T(\boldsymbol{\gamma})$. Then, we can set

$$\tilde{\alpha}_l(t) = \mathbf{B}_l^T(t)\tilde{\boldsymbol{\gamma}}_l, l = 1, \dots, p.$$

Theorem 2 establishes the theoretical property of the estimator after smoothing, which enjoys the same convergence rate.

Theorem 2 *Let Ω be any symmetric and positive definite matrix with bounded eigenvalues. Under Conditions (C1)–(C7), there exists a local minimizer of (9) such that*

$$\|\tilde{\alpha}_l - \alpha_l\|_2^2 = O_p\left(n^{-2r/(2r+1)}\right), l = 1, \dots, p.$$

With the above smoothing method, we can use the standard Newton–Raphson method to obtain the estimator. However, the closed-form derivative of $\tilde{Q}_n(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ is very messy to say the least and thus we use the two-stage method as mentioned in Greene (2011). First, the initial estimator is obtained by the R package **quantreg** with the weights $\Gamma_i, i = 1, \dots, n$ and correlations ignored. Then, the initial estimator is used to calculate Σ_n which is then fixed and standard Newton–Raphson method is used to obtain the refined estimator. According to our experience, the proposed method based on Newton–Raphson algorithm is very fast and typically converges within 20 iterations in our numerical results. Although multiple-stage approach could also be used in which refined estimator is used to further update Σ_n , we find this does not improve over the two-stage approach and thus do not use multiple-stage approach in our study.

Remark 4 In practice, the matrix Ω can be updated as an estimate of the covariance matrix of the estimators, which can be simply calculated as

$$\hat{\Omega} = \left(\{\tilde{S}'(\tilde{\boldsymbol{\gamma}})\}^T \tilde{\Sigma}_n^{-1}(\tilde{\boldsymbol{\gamma}})\tilde{S}'(\tilde{\boldsymbol{\gamma}})\right)^{-1},$$

where $\tilde{S}'(\tilde{\boldsymbol{\gamma}})$ is the derivative of $\tilde{S}(\tilde{\boldsymbol{\gamma}})$. This is the induced smoothing method proposed by Brown and Wang (2005), and Leng and Zhang (2014) adopted the same formula for longitudinal quantile linear regression. However, due to technical difficulties, we are not able to establish the validity of this rigorously due to the unsmooth nature of the objective function, and it is hard to deal with the spline approximation bias. Thus, the proposed asymptotic variance formula ignores the bias in function estimation and uses directly the formula pretending it is a linear model after spline expansion.

Remark 5 Although we are not able to provide rigorous asymptotic normality results for the nonparametric functions, we can informally argue that the QIF estimator is more efficient than the estimator assuming working independence as follows. Since the QIF

estimator is a GMM (generalized method of moments) estimator, it is well known that the asymptotic variance of the QIF estimator is at least as small as that obtained by any linear combinations of the estimating equations. When one of the basis matrices in (5) is the identity matrix corresponding to the estimating equations ignoring correlations, we have that QIF is more efficient than assuming working independence. In fact, in practice one always set one of the basis matrices to be the identity matrix.

To implement the above estimating procedure, one needs to select the order and the number of knots in spline approximation. In all simulation studies and real data analysis, for simplicity, the nonparametric functions are approximated by cubic splines and the same number of interior knots, i.e., $K := K_1 = \dots = K_p$. The number of interior knots is chosen from the interval $[n^{1/(2d+1)}, 5n^{1/(2d+1)}]$ by minimizing the following BIC criterion

$$\text{BIC}(K) = \tilde{Q}_n(\tilde{\gamma}) + p(K + d)\log(n).$$

This range for K satisfies the order assumptions in Theorem 1 on the number of interior knots K , which has also been used in Ma et al. (2014).

3 Variable selection

In practice, there are often many covariates in model (1). With high-dimensional covariates, sparse modeling is often considered superior, owing to enhanced model predictability and interpretability. In this section, we address the variable selection problem for quantile varying coefficient model based on the QIF method.

There exist some works focusing on the variable selection methods for conditional mean or conditional quantile of VCM. For example, Wang and Xia (2009) proposed the variable selection approach for VCM using kernel smoothing adaptive group LASSO (Tibshirani 1996; Yuan and Lin 2006; Zou 2006), and Zhao et al. (2013) extended the method to the quantile VCM. Noh et al. (2012) applied the polynomial spline approximation and group SCAD with local linear approximation (LLA, Zou and Li 2008) to investigate variable selection for quantile VCM. Verhasselt (2014) discussed the variable selection for the generalized VCM using P-splines. For longitudinal data, Wang et al. (2008) studied the variable selection approach via group SCAD penalty (Fan and Li 2001). However, it ignored the correlation within subjects, which may lead to some efficiency loss in estimation and variable selection. Here, we consider variable selection of quantile VCM that incorporates the correlations within subjects.

To conduct variable selection, we propose the penalized estimator by minimizing the following penalized QIF, defined as

$$\hat{\gamma}^P = \operatorname{argmin}_{\gamma} \left\{ Q_n(\gamma) + n \sum_{l=1}^p p_{\lambda_l} (\|\gamma_l\|_{H_l}) \right\}, \tag{10}$$

where $p_{\lambda_l}(\cdot)$ is a given penalty function depending on the tuning parameter λ_l , $l = 1, \dots, p$, and $\|\gamma_l\|_{H_l}^2 = \gamma_l^T H_l \gamma_l$ with $H_l = \int B_l(t) B_l^T(t) dt$. Notice that $\|\gamma_l\|_{H_l} =$

$\|s_l\|_2$, the L_2 norm of the spline function $s_l(\cdot)$. Shrinking $\|s_l\|_2$ to 0 entails $s_l \equiv 0$. In addition, the tuning parameters λ_l for the penalty functions in (10) are not necessarily the same for different l , which can provide further flexibility.

There are several possible choices for the penalty function $p_{\lambda_l}(\cdot)$, such as LASSO (Tibshirani 1996), MCP (Zhang 2010) and SCAD penalty (Fan and Li 2001). Here, we choose the SCAD penalty, which is defined as

$$p_\lambda(\theta) = \begin{cases} \lambda|\theta|, & |\theta| \leq \lambda, \\ -(\theta^2 - 2a\lambda|\theta| + \lambda^2)/[2(a - 1)], & \lambda < |\theta| \leq a\lambda, \\ (a + 1)\lambda^2/2, & |\theta| > a\lambda, \end{cases}$$

for given $a > 2$ and $\lambda > 0$. The SCAD penalty is continuously differentiable on $(-\infty, 0) \cup (0, \infty)$ but singular at 0, and its derivative vanishes outside $[-a\lambda, a\lambda]$, which can produce sparse estimates for small coefficients and unbiased estimates for large coefficients. Therefore, we obtain the penalized estimator of $\alpha_l(t)$ as

$$\hat{\alpha}_l^P(t) = \mathbf{B}_l^T(t)\hat{\boldsymbol{\gamma}}_l^P, \quad l = 1, \dots, p.$$

We next discuss the asymptotic properties of the resulting penalized estimator. Without loss of generality, we assume that $\alpha_l(\cdot) = 0, j = d_0 + 1, \dots, p$ and $\alpha_l(\cdot), l = 1, \dots, d_0$ are all nonzero components of $\boldsymbol{\alpha}(\cdot)$. We first show in Theorem 3 that the penalized QIF estimators $\{\hat{\alpha}_l^P(t)\}_{l=1}^p$ have the same convergence rate as the unpenalized estimators $\{\hat{\alpha}_l(t)\}_{l=1}^p$. Moreover, Theorem 4 establishes the sparsity property of the penalized estimators, that is, $\hat{\alpha}_l^P(t) = 0$ with probabilities approaching one for $l = d_0 + 1, \dots, p$.

Theorem 3 *Under the conditions of Theorem 1, if the tuning parameters satisfy $\max_l \lambda_l \rightarrow 0$ in probability as $n \rightarrow \infty$, then there exists a local minimizer of (10) such that*

$$\|\hat{\alpha}_l^P - \alpha_l\|_2^2 = O_p\left(n^{-2r/(2r+1)}\right), \quad l = 1, \dots, p.$$

Theorem 4 *Under the same conditions of Theorem 3, if the tuning parameters further satisfy $\min_{d_0+1 \leq l \leq p} \lambda_l n^{r/(2r+1)} \rightarrow \infty$ in probability as $n \rightarrow \infty$, then, with probability approaching 1, $\hat{\alpha}_l^P(t) = 0$ for $l = d_0 + 1, \dots, p$.*

Since the penalty function is not differentiable, to obtain the penalized estimator $\hat{\boldsymbol{\gamma}}^P$, we need to smooth the penalized object function for both terms in (10). For $\mathbf{Q}_n(\cdot)$, we can use the induced smoothing method as in Sect. 2.3, and for penalty $p_{\lambda_l}(\cdot)$, we can use the local quadratic approximation (LQA, Fan and Li 2001). Then, we can use the Newton–Raphson algorithm to obtain the penalized smoothing QIF estimator, denoted by $\tilde{\boldsymbol{\gamma}}^P$. As in Theorem 2, we can establish that the penalized estimator $\tilde{\boldsymbol{\gamma}}^P$ still enjoys the same asymptotic property as $\hat{\boldsymbol{\gamma}}^P$. The details are omitted.

Unlike our approach, Noh et al. (2012) used the local linear approximation (LLA) to deal with the penalty terms, which allows the authors to convert the optimization problem to a second-order cone programming. We just use the simple LQA approach in

order to make the optimization problem differentiable. Our model is more complicated, and it remained to be seen whether LLA will allow a more efficient algorithm to be developed.

Note that there are p tuning parameters to be chosen in conducting the variable selection procedure. To reduce the computational burden, we propose to use the following strategy by setting

$$\lambda_l = \frac{\lambda}{\|\tilde{\boldsymbol{y}}_l\|_{H_l}}, \quad l = 1, \dots, p,$$

where $\tilde{\boldsymbol{y}}_l$ is the initial unpenalized estimator of \boldsymbol{y}_l obtained in Sect. 2.3. Note that the above strategy has also been used in Wang and Xia (2009). Then, we use the following criterion to obtain the optimal tuning parameter

$$\hat{\lambda}^{\text{opt}} = \operatorname{argmin}_{\lambda} \left\{ \tilde{\boldsymbol{Q}}_n(\tilde{\boldsymbol{y}}^P) + J \cdot \log(n) \cdot df_{\lambda} \right\},$$

where df_{λ} is the number of nonzero coefficient functions for a given tuning parameter λ .

4 Simulation studies

In this section, we conduct simulation studies to illustrate our proposed methods. Note that the minimizer of \boldsymbol{Q}_n is presented merely for theoretical reasons and we are not able to compute it due to the discontinuity of the objective function, while the minimizer of $\tilde{\boldsymbol{Q}}_n$ can be easily found using the two-stage approach as mentioned in Sect. 2. For each example, we focus on $\tau = 0.25$ and 0.5 , and 500 data sets are generated to evaluate the simulation results.

Example 1 In this example, the responses y_{ij} are generated from

$$y_{ij} = \alpha_0(t_{ij}) + x_{ij1}\alpha_1(t_{ij}) + x_{ij2}\alpha_2(t_{ij}) + (1 + \kappa|t_{ij}|)\epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i,$$

where the number of subjects is $n = 50, 100$ and 150 , and each subject is supposed to be measured at scheduled time point t_{ij} from $\{0, 0.1, 0.2, \dots, 1\}$, each of which has a 20% probability of being skipped. The actual measurement times are generated by adding a $U(-0.05, 0.05)$ random variable to the scheduled times. The three nonparametric functions are set to be

$$\alpha_0(t) = 3 \sin(2\pi t), \quad \alpha_1(t) = 8t(1 - t) \quad \text{and} \quad \alpha_2(t) = 2 \cos(2\pi t).$$

The marginal distributions of the two covariates are standard normal, and their correlation coefficient is 0.5. The random error $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$ follows multivariate normal distribution or multivariate t -distribution (degrees of freedom 3) with location parameter $-q_{\tau}$ and covariance matrix Σ , where q_{τ} is the τ -th quantile of the standard

Table 1 Simulation results of AIMSE when the true error structure is compound symmetry

Error	Model	τ	n	WI	QIF-CS	QIF-AR(1)		
Normal	HM	0.25	100	0.0385 (0.0148)	0.0332 (0.0151)	0.0380 (0.0160)		
			150	0.0287 (0.0108)	0.0231 (0.0099)	0.0279 (0.0123)		
			200	0.0224 (0.0071)	0.0182 (0.0073)	0.0208 (0.0070)		
		0.5	100	0.0334 (0.0123)	0.0240 (0.0105)	0.0276 (0.0106)		
			150	0.0251 (0.0087)	0.0171 (0.0055)	0.0195 (0.0060)		
			200	0.0201 (0.0064)	0.0143 (0.0040)	0.0157 (0.0046)		
	HT	0.25	100	0.0599 (0.0316)	0.0369 (0.0174)	0.0507 (0.0246)		
			150	0.0406 (0.0171)	0.0326 (0.0142)	0.0406 (0.0204)		
			200	0.0311 (0.0129)	0.0260 (0.0112)	0.0299 (0.0124)		
		0.5	100	0.0406 (0.0176)	0.0277 (0.0113)	0.0336 (0.0137)		
			150	0.0290 (0.0111)	0.0202 (0.0078)	0.0229 (0.0076)		
			200	0.0235 (0.0082)	0.0167 (0.0055)	0.0184 (0.0058)		
		t	HM	0.25	100	0.0820 (0.0362)	0.0480 (0.0283)	0.0684 (0.0341)
					150	0.0578 (0.0267)	0.0383 (0.0143)	0.0540 (0.0263)
					200	0.0441 (0.0177)	0.0290 (0.0109)	0.0389 (0.0161)
0.5	100			0.0682 (0.0283)	0.0439 (0.0200)	0.0557 (0.0241)		
	150			0.0501 (0.0216)	0.0305 (0.0116)	0.0372 (0.0147)		
	200			0.0385 (0.0149)	0.0239 (0.0078)	0.0284 (0.0101)		
HT	0.25		100	0.1317 (0.0730)	0.0638 (0.0288)	0.1024 (0.0536)		
			150	0.0852 (0.0395)	0.0547 (0.0237)	0.0834 (0.0453)		
			200	0.0645 (0.0297)	0.0431 (0.0187)	0.0607 (0.0289)		
	0.5		100	0.0861 (0.0423)	0.0531 (0.0237)	0.0687 (0.0312)		
			150	0.0597 (0.0263)	0.0363 (0.0147)	0.0442 (0.0174)		
			200	0.0464 (0.0192)	0.0301 (0.0116)	0.0348 (0.0142)		

The values in parentheses are the standard errors of the AIMSE over the 500 replications

normal distribution or t -distribution with degrees of freedom 3, which implies the corresponding τ -th quantile of ϵ_{ij} is zero. The covariance matrix Σ follows either the compound symmetry (CS) or AR(1) structure with parameter $\rho = 0.8$. In addition, the quantity κ equals 0 or 1 corresponding to homoscedastic model (HM) and heteroscedastic model (HT), respectively.

To assess the estimation efficiency for nonparametric functions, we calculate the integrated mean square error (IMSE) defined as

$$IMSE(\alpha_l) = \frac{1}{n_{grid}} \sum_{k=1}^{n_{grid}} \{\alpha_l(t_k) - \hat{\alpha}_l(t_k)\}^2$$

averaged over 500 data sets and report the average of the integrated mean square error (AIMSE)

$$AIMSE = \frac{1}{P} \sum_{l=1}^P IMSE(\alpha_l),$$

Table 2 Simulation results of AIMSE when the true error structure is AR(1)

Error	Model	τ	n	WI	CS	AR(1)
Normal	HM	0.25	100	0.0147 (0.0054)	0.0145 (0.0057)	0.0142 (0.0055)
			150	0.0104 (0.0035)	0.0097 (0.0034)	0.0091 (0.0031)
			200	0.0083 (0.0026)	0.0076 (0.0024)	0.0070 (0.0022)
		0.5	100	0.0126 (0.0044)	0.0118 (0.0040)	0.0102 (0.0036)
			150	0.0090 (0.0029)	0.0080 (0.0024)	0.0070 (0.0022)
			200	0.0074 (0.0023)	0.0064 (0.0019)	0.0058 (0.0017)
	HT	0.25	100	0.0315 (0.0133)	0.0275 (0.0128)	0.0258 (0.0119)
			150	0.0213 (0.0087)	0.0175 (0.0069)	0.0155 (0.0061)
			200	0.0165 (0.0065)	0.0133 (0.0052)	0.0115 (0.0046)
		0.5	100	0.0790 (0.0332)	0.0732 (0.0312)	0.0638 (0.0281)
			150	0.0549 (0.0214)	0.0476 (0.0175)	0.0411 (0.0161)
			200	0.0431 (0.0171)	0.0370 (0.0138)	0.0324 (0.0127)
t	HM	0.25	100	0.0678 (0.0320)	0.0608 (0.0285)	0.0567 (0.0276)
			150	0.0447 (0.0167)	0.0418 (0.0163)	0.0400 (0.0163)
			200	0.0353 (0.0125)	0.0321 (0.0116)	0.0307 (0.0112)
		0.5	100	0.0456 (0.0164)	0.0423 (0.0155)	0.0372 (0.0147)
			150	0.0316 (0.0111)	0.0286 (0.0100)	0.0250 (0.0083)
			200	0.0253 (0.0084)	0.0224 (0.0070)	0.0202 (0.0066)
	HT	0.25	100	0.1478 (0.0742)	0.1297 (0.0641)	0.1230 (0.0636)
			150	0.0959 (0.0423)	0.0774 (0.0330)	0.0699 (0.0335)
			200	0.0746 (0.0318)	0.0599 (0.0261)	0.0530 (0.0239)
		0.5	100	0.0989 (0.0418)	0.0931 (0.0429)	0.0805 (0.0373)
			150	0.0649 (0.0270)	0.0592 (0.0254)	0.0508 (0.0214)
			200	0.0507 (0.0209)	0.0453 (0.0181)	0.0395 (0.0163)

where $\{t_k : k = 1, \dots, n_{\text{grid}}\}$ with $n_{\text{grid}} = 200$ are the grid time points at which the functions $\{\alpha_l(\cdot)\}$ are evaluated. The simulation results are shown in Tables 1 and 2, where we also report the corresponding results by assuming working independence (WI) for comparison.

Table 1 summarizes the estimation results when the error correlation has compound symmetry structure. The AIMSEs for each method become smaller as the sample size increases. Moreover, it shows that the estimators with a correct CS working correlation have the smallest AIMSE, and even with misspecified AR(1) working correlation the efficiency gains are also obvious compared with WI ignoring the correlation within subjects. If the model is heteroscedastic and/or the error follows multivariate t -distribution, the efficiency gain is more obvious. Similar phenomena are also observed for the case of the true error correlation being AR(1) as shown in Table 2. Further simulation studies (not shown) indicate that the performances when incorporating $\mathbf{\Gamma}_i$ are only a little better than replacing $\mathbf{\Gamma}_i$ with an identity matrix. In addition, we have also tried the simulations when the error has lower correlation $\rho = 0.3$ or $\rho = 0.5$; the performances of our proposed approach are better than or at least as well as WI. We omit them to save space.

Example 2 In this example, the data are generated from the following model:

$$y_{ij} = \sum_{l=1}^p x_{ijl} \alpha_l(t_{ij}) + (1 + \kappa |x_{ij3}|) \epsilon_{ij}, \quad i = 1, \dots, 250, \quad j = 1, \dots, m_i.$$

Each subject is supposed to be measured at scheduled time points $\{0, 1, 2, \dots, 20\}$, and each time point has a 50% probability of being skipped. Similar to Example 1, the actual measurement times are generated by adding a $U(-0.5, 0.5)$ random variable to the scheduled times. The three relevant variables, x_{ijl} , $l = 1, 2, 3$, are simulated as follows: x_{ij1} is generated from $U(t_{ij}/10, 2 + t_{ij}/10)$, x_{ij2} , conditional on x_{ij1} , is $N(0, (1 + x_{ij1})/(2 + x_{ij1}))$, and x_{ij3} , independent of x_{ij1} and x_{ij2} , is a Bernoulli random variable with success probability 0.8. The rest five redundant variables, x_{ijl} , $l = 4, \dots, p$, are mutually independent, and for each l , x_{ijl} is generated from a multivariate Gaussian distribution with zero mean and a decayed exponential covariance

$$\text{cov}(x_{ijl}, x_{iuv}) = \begin{cases} 4 \exp(-|t_{ij} - t_{uv}|), & \text{if } i = u \\ 0, & \text{if } i \neq u \end{cases}$$

The three nonparametric functions are

$$\alpha_1(t) = 15 + 20 \sin\left(\frac{\pi t}{40}\right), \quad \alpha_2(t) = 2 - 3 \cos\left(\frac{(t - 25)\pi}{15}\right), \quad \alpha_3(t) = 6 - 0.6t.$$

The marginal variance of random error $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{im_i})^T$ is 4, and the correlation settings are the same as in Example 1, and the cases $\kappa = 0$ and 1 are still denoted by HM and HT, respectively.

We report the results with $p = 8$ in Tables 3 and 4, where the simulation results obtained by ignoring the correlation are also shown. Moreover, we also report the results with group LASSO penalty in Tables 3 and 4 for comparisons. In Table 3, the oracle estimator is obtained using only the first three relevant variables. In Table 4, two quantities “False positive rate” (FPR) and “false negative rate” (FNR) are used to evaluate the performance of variable selection, where FNR denotes the proportion of nonzero varying coefficient functions incorrectly estimated as zero coefficients, while FPR denotes the proportion of zero coefficients incorrectly estimated by nonzero functions, and we present the mean proportion over 500 replications. In addition, according to the suggestion of a reviewer, we also report the simulation results for case of divergent dimension $p = \lceil n^{1/2} \rceil$ in Appendix in Supplementary Material (Tables 1 and 2).

The results in Tables 3 and 4 indicate that the performances of our proposed variable selection approach are satisfactory. Both the nonzero components and zero components can be correctly identified in terms of FPR and FNR. For variable selection, the difference between the proposed QIF method and the method assuming working independence is very small for both homoscedastic model and heteroscedastic model with group SCAD penalty or group LASSO penalty. Moreover, as we can see from

Table 3 Mean and standard deviation of AIMSE ($\times 100$) over 500 replications for Example 2 with $p = 8$

Model	Error	τ	Group SCAD penalty			Group LASSO penalty			Oracle estimate		
			WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR
HM	Normal-CS	0.25	3.3373 (1.5309)	1.6105 (0.6883)	2.0073 (0.9098)	3.5568 (1.6803)	2.5041 (1.0207)	3.3360 (1.5175)	3.3189 (1.5224)	1.5908 (0.7147)	1.9811 (0.8837)
		0.5	2.8034 (1.3100)	1.1853 (0.5531)	1.5181 (0.7172)	2.8905 (1.4434)	1.6162 (0.7338)	2.1524 (0.9962)	2.7955 (1.3255)	1.1753 (0.5468)	1.5074 (0.7081)
t-CS	Normal-AR	0.25	3.2354 (1.5077)	2.7704 (1.2892)	2.3256 (1.1271)	3.5949 (1.5716)	3.9978 (1.7554)	3.4668 (1.5272)	3.2271 (1.4728)	2.7662 (1.3059)	2.2830 (1.1130)
		0.5	2.8587 (1.2333)	2.3062 (0.9623)	1.8444 (0.8035)	3.0482 (1.6460)	2.8847 (1.6352)	2.6601 (1.3165)	2.8576 (1.2313)	2.3002 (0.9623)	1.8418 (0.8010)
t-AR	t-CS	0.25	5.1381 (2.6069)	2.3219 (1.1356)	3.0565 (1.6387)	5.3763 (2.4565)	3.6931 (1.6470)	4.8726 (2.3281)	5.1253 (2.7015)	2.2718 (1.1383)	2.9370 (1.5269)
		0.5	3.4738 (1.6466)	1.4386 (0.7320)	1.8826 (0.8919)	3.7316 (1.8974)	2.0218 (1.0093)	2.8212 (1.4456)	3.4614 (1.6483)	1.4336 (0.7309)	1.8733 (0.8970)
t-AR	t-AR	0.25	5.4807 (2.7373)	4.2977 (2.1446)	3.5271 (1.6603)	5.7803 (2.8818)	5.4316 (2.9325)	5.3116 (2.8708)	5.4602 (2.7660)	4.2436 (2.1663)	3.4351 (1.5260)
		0.5	3.5097 (1.7752)	2.8361 (1.2912)	2.2351 (1.1495)	3.7818 (1.8979)	3.5671 (1.8373)	3.0737 (1.5608)	3.4990 (1.7833)	2.8028 (1.3090)	2.2256 (1.1487)

Table 3 continued

Model	Error	τ	Group SCAD penalty			Group LASSO penalty			Oracle estimate		
			WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR
HT	Normal-CS	0.25	6.9411	3.7207	5.1324	8.7181	5.8483	8.4594	6.9001	3.6724	4.9666
			(2.9939)	(1.5393)	(2.0605)	(3.9221)	(2.4988)	(3.5840)	(3.1319)	(1.5211)	(1.9175)
			5.7342	2.6913	3.5575	6.6619	3.8525	5.5933	5.7128	2.6872	3.5545
	Normal-AR	0.25	(2.6416)	(1.1901)	(1.6569)	(3.2684)	(1.7529)	(2.4727)	(2.6705)	(1.1847)	(1.6574)
			7.1009	6.6087	5.8266	8.9307	8.7328	8.4114	7.0876	6.5902	5.7678
			(3.0169)	(3.0276)	(2.5509)	(3.7070)	(3.2583)	(3.0207)	(3.0292)	(3.0220)	(2.5390)
	t-CS	0.5	6.1418	5.3626	4.6134	7.2872	7.1946	6.7742	6.1290	5.3588	4.6054
			(2.6277)	(2.1281)	(1.9689)	(3.7790)	(3.5731)	(3.1193)	(2.6559)	(2.1314)	(1.9663)
			11.1576	5.7204	7.8611	12.3867	8.4729	8.9033	11.0639	5.5393	7.6074
	t-AR	0.5	(5.2111)	(2.6912)	(3.7333)	(5.1856)	(4.0804)	(4.8128)	(5.4844)	(2.5367)	(3.6707)
			7.6037	3.4608	4.7681	8.6386	4.8957	6.9101	7.5519	3.4383	4.7723
			(3.6560)	(1.6089)	(2.4945)	(4.3589)	(2.4334)	(3.7069)	(3.8856)	(1.5892)	(2.5029)
	t-AR	0.25	11.6760	11.1541	10.4610	14.9257	14.3101	14.1287	11.0163	10.1791	8.7519
			(5.4772)	(5.5648)	(4.9787)	(8.0563)	(6.8486)	(6.6055)	(5.6100)	(4.8745)	(3.7826)
			7.4309	6.8767	5.7767	8.9244	8.0865	7.7187	7.4804	6.7633	5.6419
		0.5	(3.2387)	(2.8200)	(2.5164)	(4.0756)	(4.0387)	(3.3210)	(3.2016)	(2.8298)	(2.5194)

Table 4 Variable selection results for Example 2 with $p = 8$

Penalty	Model	Error	τ	FNR			FPR		
				WI	QIF-CS	QIF-AR	WI	QIF-CS	QIF-AR
Group SCAD	HM	Normal-CS	0.25	0	0	0	0.002	0.003	0.003
			0.5	0	0	0	0.001	0.001	0.005
		Normal-AR	0.25	0	0	0	0.003	0.005	0.003
			0.5	0	0	0	0.002	0.004	0.002
		t-CS	0.25	0	0	0	0.005	0.006	0.008
			0.5	0	0	0	0.004	0.004	0.005
	t-AR	0.25	0	0	0	0.008	0.007	0.006	
		0.5	0	0	0	0.003	0.005	0.005	
	HT	Normal-CS	0.25	0	0	0	0.013	0.012	0.013
			0.5	0	0	0	0.011	0.010	0.010
		Normal-AR	0.25	0	0	0	0.011	0.009	0.007
			0.5	0	0	0	0.008	0.007	0.005
		t-CS	0.25	0	0	0	0.021	0.023	0.026
			0.5	0	0	0	0.018	0.019	0.018
t-AR		0.25	0	0	0	0.024	0.022	0.024	
		0.5	0	0	0	0.013	0.012	0.011	
Group LASSO	HM	Normal-CS	0.25	0	0	0	0.004	0.003	0.005
			0.5	0	0	0	0.004	0.004	0.006
		Normal-AR	0.25	0	0	0	0.007	0.006	0.006
			0.5	0	0	0	0.004	0.005	0.005
		t-CS	0.25	0	0	0	0.007	0.007	0.009
			0.5	0	0	0	0.005	0.005	0.006
	t-AR	0.25	0	0	0	0.010	0.011	0.010	
		0.5	0	0	0	0.006	0.008	0.007	
	HT	Normal-CS	0.25	0	0	0	0.015	0.016	0.018
			0.5	0	0	0	0.011	0.010	0.011
		Normal-AR	0.25	0	0	0	0.018	0.017	0.016
			0.5	0	0	0	0.013	0.011	0.012
		t-CS	0.25	0	0	0	0.027	0.033	0.035
			0.5	0	0	0	0.024	0.022	0.021
t-AR		0.25	0	0	0	0.035	0.032	0.034	
		0.5	0	0	0	0.023	0.025	0.022	

Table 3, AIMSEs calculated based on SCAD penalized estimates are closer to the AIMSEs from the oracle estimator than the group LASSO penalty. In addition, the QIF-based estimation performance is a little better than working independence even if we use the misspecified working correlation matrix. This shows that our penalized QIF estimators can simultaneously estimate and select important variables and gain estimation accuracy by effectively removing the zero component variables. The

above findings confirm our theoretical results and demonstrate efficiency gains by our proposed approach compared to ignoring correlations within subjects.

For the case of divergent dimension $p = \lceil n^{1/2} \rceil$, the proportion of FNR in Table 2 in Supplementary Material is still very small while the proportion of FPR for heteroscedastic model is noticeably larger. However, the gains of our method are still obvious in terms of AIMSEs in Table 1.

5 Real data analysis

In this section, we demonstrate an application of our proposed method to a longitudinal AIDS data from the Multi-Center AIDS Cohort Study between 1984 and 1991. This AIDS Cohort Study was firstly reported in [Kaslow et al. \(1987\)](#), where each individual was scheduled to undergo measurements at semiannual visits, but because many individuals missed some of their scheduled visits and the human immunodeficiency virus (HIV) infections occurred randomly during the study, there were unequal numbers of repeated measurements and different measurement times for each individual. As a subset of the cohort, our analysis focused on the 283 homosexual men who were infected with HIV during the follow-up period. The main interest of these data is to describe the trend of the level of the CD4 percentage depletion over time and to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage and age at infection on the CD4 percentage after the infection. These data have been studied in several papers, including [Huang et al. \(2002\)](#), [Qu and Li \(2006\)](#), [Fan et al. \(2007\)](#) and [Wang et al. \(2009\)](#). Recently, [Wang et al. \(2008\)](#) have considered variable selection in varying coefficients models for these data based on mean regression and quantile regression, respectively. However, they did not consider the correlation between measurements for the same individual, which may lose estimation efficiency.

In the following, our analysis focuses on evaluating the time-dependent effects of smoking status (x_{i1} , taking values of 1 and 0 whether smoke or not), age (x_{i2}), preCD4 (x_{i3} , pre-HIV infection CD4 percentage) and the interaction of the covariates at different quantile levels. We consider the following varying coefficient model:

$$y_{ij} = \alpha_0(t_{ij}, \tau) + \alpha_1(t_{ij}, \tau)x_{i1} + \alpha_2(t_{ij}, \tau)x_{i2} + \alpha_3(t_{ij}, \tau)x_{i3} + \alpha_4(t_{ij}, \tau)x_{i2}^2 \\ + \alpha_5(t_{ij}, \tau)x_{i3}^2 + \alpha_6(t_{ij}, \tau)x_{i1}x_{i2} + \alpha_7(t_{ij}, \tau)x_{i1}x_{i3} + \alpha_8(t_{ij}, \tau)x_{i2}x_{i3} + \epsilon_{ij}(\tau),$$

where y_{ij} is the i -th individual's CD4 percentage at time t_{ij} (in years), and x_{i2} and x_{i3} are standardized. The baseline function $\alpha_0(t, \tau)$ represents the τ -th quantile of CD4 percentage t years after the infection for a nonsmoker with average preCD4 percentage and average age at HIV infection. Here, we focus on three quantile levels $\tau = 0.25, 0.5$ and 0.75 , and we use the compound symmetry working correlation to fit the data, and the results are compared with working independence. The results of AR(1) working correlation are very similar.

Table 5 shows the selected components at different quantile levels, and we found that our QIF method selects one or two more variables than WI approach at quantile levels $\tau = 0.25$ and 0.5 . In particular, the smoking effect is selected at $\tau = 0.25$ but not at $\tau = 0.5$ and 0.75 , which indicates that smoking has the effect on the CD4 counts when

Table 5 Selected components of two different methods at three quantile levels for AIDS data

τ	QIF-CS	WI
0.25	$\alpha_0, \alpha_1, \alpha_3, \alpha_6$	α_0, α_3
0.5	$\alpha_0, \alpha_3, \alpha_6, \alpha_7$	$\alpha_0, \alpha_3, \alpha_6$
0.75	$\alpha_0, \alpha_3, \alpha_6, \alpha_7$	$\alpha_0, \alpha_3, \alpha_6, \alpha_7$

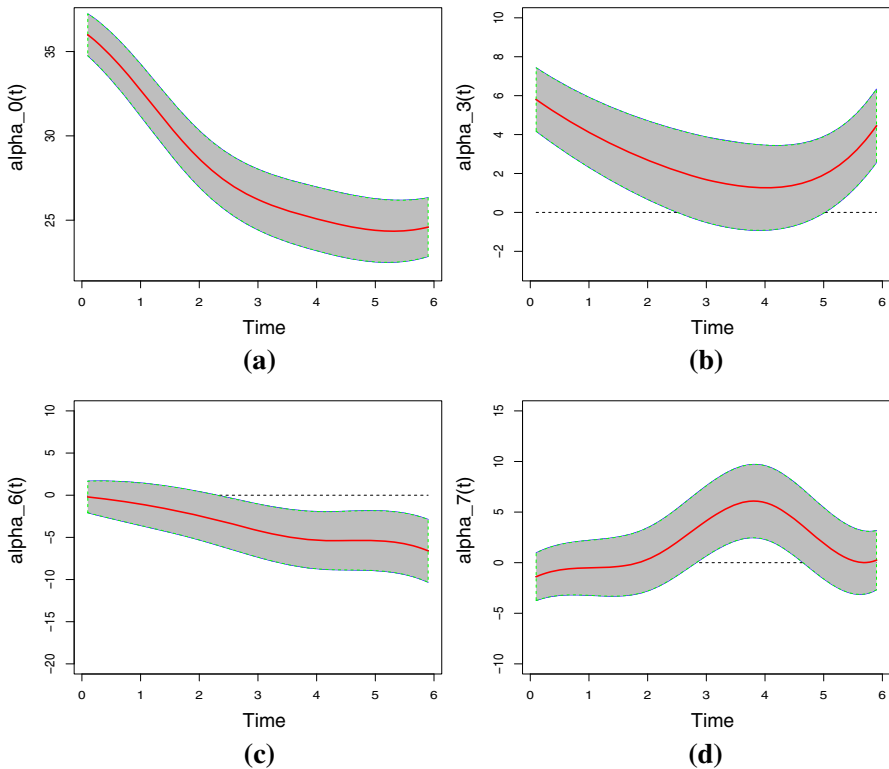


Fig. 1 Estimated coefficient curves at $\tau = 0.5$. **a** The baseline coefficient function, **b** coefficient for preCD4, **c** coefficient for the interaction of smoking and age and **d** coefficient for the interaction of smoking and preCD4. The shaded area indicates the 95% point-wise confidence band

the count is small. The estimated curves at $\tau = 0.5$ together with their 95% point-wise confidence band of the four important nonparametric components are shown in Fig. 1. Note that the zero lines are not completely contained in the selected 95% point-wise confidence bands, which indicate that the variables selected are reasonable. Figure 1a implies that the baseline coefficient function $\alpha_0(t, \tau)$ decreases over time. From Fig. 1b, the preCD4 has positive effect on the post-infection CD4 on the whole with the rate decreases at first and increases later. It is noteworthy that the interaction between smoking and age is significant from Fig. 1c, though the individual variables have no effects on the post-infection CD4. In particular, the elder smokers tend to have lower CD4 counts. Furthermore, Fig. 1d shows that interaction between smoking and

preCD4 increases at first and then decreases, which implies that smokers with high preCD4 tend to have higher CD4 counts after infection and it decreases after a period of follow-up. One possible explanation is that a smoking patient with lower preCD4 may choose to quit smoking due to medical concerns at first, and they start to smoke again when their preCD4 counts improved after a period.

6 Concluding remarks

In this paper, we investigate statistical method and theory for estimation and variable selection of quantile varying coefficient model in longitudinal data analysis. By using the QIF approach, the correlation within subject is incorporated and the efficiency of estimation and accuracy of the variable selection are improved compared to quantile regression assuming working independence. We further propose to use induced smoothing for estimation, and the procedure is easy to implement. Simulation studies and real data analysis show that the performance of our proposed approach is encouraging.

Several problems can be investigated in the future. Note that the dimension of the nonparametric components is fixed in this work. It may be extended to the case with a diverging number of covariates. One can also extend this work to the partially linear varying coefficient model, which is more flexible than VCM. In addition, it is also important to identify varying and constant coefficients among the nonzero coefficients. For that purpose, it is possible to use the hypotheses testing approach like Wang et al. (2009) after identifying the nonzero coefficients based on our method. Finally, since quantile regression curves are estimated individually, one interesting and important problem is how to avoid the crossing of the estimated quantile curves at adjacent quantile levels (Bondell et al. 2010).

Acknowledgements We sincerely thank the Editor/Professor Hironori Fujisawa, an associate editor and two anonymous reviewers, for their insightful comments that have led to significant improvement of the paper. Zhao's research is supported in part by National Social Science Foundation of China (15BTJ027). Zhang's research is supported by the National Natural Science Foundation of China Grant 11671374 and 71631006. Heng Lian's research is partially supported by City University of Hong Kong Start-up Grant 7200521 and RGC General Research Fund 11301718.

References

- Andriyana, Y., Gijbels, I., Verhasselt, A. (2014). P-splines quantile regression estimation in varying coefficient models. *Test*, 23(1), 153–194.
- Bondell, H. D., Reich, B. J., Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4), 825–838.
- Brown, B., Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, 92(1), 149–158.
- Cai, Z., Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484), 371–383.
- Chiang, C.-T., Rice, J. A., Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96(454), 605–619.

- De Boor, C. (2001). *A practical guide to splines* (revised ed., Vol. 27.), Applied mathematical sciences New York: Springer.
- Fan, J., Huang, T., Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102(478), 632–641.
- Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.
- Fan, J., Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 27, 1491–1518.
- Fan, J., Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27(4), 715–731.
- Fu, L., Wang, Y.-G. (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8), 2526–2538.
- Greene, W. H. (2011). *Econometric Analysis*. New York, USA: Pearson.
- Hall, P., Sheather, S. J. (1988). On the distribution of a studentized quantile, *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3), 381–391.
- Hastie, T., Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B (Methodological)*, 55(4), 757–796.
- He, X., Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, 58(2), 162–181.
- Hendricks, W., Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87(417), 58–68.
- Huang, J. Z., Wu, C. O., Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1), 111–128.
- Huang, J. Z., Zhang, L., Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustering data using splines. *Scandinavian Journal of Statistics*, 34(3), 451–477.
- Jung, S.-H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91(433), 251–257.
- Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., Rinaldo, C. R., et al. (1987). The multicenter aids cohort study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126(2), 310–318.
- Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1), 92–108.
- Koenker, R. (2005). *Quantile regression*. Cambridge, UK: Cambridge University Press.
- Leng, C., Zhang, W. (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing*, 24(1), 123–136.
- Li, G., Lai, P., Lian, H. (2014). Variable selection and estimation for partially linear single-index models with longitudinal data, *Statistics and Computing* (in press).
- Lian, H., Liang, H., Wang, L. (2014). Generalized additive partial linear models for clustered data with diverging number of covariates using gee. *Statistica Sinica*, 24(1), 173–196.
- Liang, K.-Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.
- Lin, H., Song, P. X.-K., Zhou, Q. M. (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhyā: The Indian Journal of Statistics*, 69(3), 581–614.
- Ma, S., Liang, H., Tsai, C.-L. (2014). Partially linear single index models for repeated measurements. *Journal of Multivariate Analysis*, 130, 354–375.
- Noh, H., Chung, K., Van Keilegom, I., et al. (2012). Variable selection of varying coefficient models in quantile regression. *Electronic Journal of Statistics*, 6, 1220–1238.
- Qu, A., Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62(2), 379–391.
- Qu, A., Lindsay, B. G., Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87(4), 823–836.
- Şentürk, D., Müller, H.-G. (2008). Generalized varying coefficient models for longitudinal data. *Biometrika*, 95(3), 653–666.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge, UK: Cambridge University Press.
- Verhasselt, A. (2014). Generalized varying coefficient models: a smooth variable selection technique. *Statistica Sinica*, 24(1), 147–171.

- Wang, H. J., Zhu, Z., Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, 37(6), 3841–3866.
- Wang, H., Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104(486), 747–757.
- Wang, L., Li, H., Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, 103(484), 1556–1569.
- Wang, N., Carroll, R. J., Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100(469), 147–157.
- Xue, L., Qu, A., Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association*, 105(492), 1518–1530.
- Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.
- Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894–942.
- Zhao, W., Zhang, R., Lv, Y., Liu, J. (2013). Variable selection of the quantile varying coefficient regression models. *Journal of the Korean Statistical Society*, 42(3), 343–358.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418–1429.
- Zou, H., Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of statistics*, 36(4), 1509–1533.