Journal of Statistical Computation and Simulation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Robust and smoothing variable selection for quantile regression models with longitudinal data

Z. C. Fu, L. Y. Fu & Y. N. Song

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2023, VOL. 93, NO. 15, 2600–2624
https://doi.org/10.1080/00949655.2023.2201007

Robust and smoothing variable selection for quantile regression models with longitudinal data

Z. C. Fu[a,b], L. Y. Fu[a] and Y. N. Song[a]

[a]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, People's Republic of China;
[b]Department of Industrial Engineering, Center for Statistical Science, Tsinghua University, Beijing, People's Republic of China

ABSTRACT

In this paper, we propose a penalized weighted quantile estimating equations (PWQEEs) method to obtain sparse, robust and efficient estimators for the quantile regression with longitudinal data. The PWQEE incorporates the within correlations in the longitudinal data by Gaussian copulas and can also down-weight the high leverage points in covariates to achieve double-robustness to both the non-normal distributed errors and the contaminated covariates. To overcome the obstacles of discontinuity of the PWQEE and nonconvex optimization, a local distribution smoothing method and the minimization–maximization algorithm are proposed. The asymptotic properties of the proposed method are also proved. Furthermore, finite sample performance of the PWQEE is illustrated by simulation studies and a real-data example.

ARTICLE HISTORY
Received 21 April 2022
Accepted 2 April 2023

KEYWORDS
Correlation matrix; generalized estimating equations; robust; variable selection

1. Introduction

Variable selection plays an important role in high-dimensional data analysis. In the model-building process, it is common to have a large number of candidate predictors. Variable selection techniques are often adopted to identify and exclude the irrelevant explanatory variables to make the models easier to interpret and improve the predictive ability. In the regularization framework, many penalties have been proposed to obtain sparse estimators. Tibshirani [1] proposed the Lasso method which uses the L_1 penalty, $p_\lambda(|\beta|) = \lambda|\beta|$. Bridge regression [2] adopted the L_q penalty, $p_\lambda(|\beta|) = \lambda|\beta|^q$, to achieve variable selection. When $0 < q < 1$, the L_q penalty is a nonconvex function on $(0, \infty)$. Then, Fan and Li [3] proposed a smoothly clipped absolute deviation (SCAD) penalty and gave the desirable properties of a good penalty which were called 'oracle' properties. Both the Lasso and the bridge regression did not satisfy the 'oracle' properties. Zou [4] proposed an adaptive-Lasso method which is an extension of Lasso and has the oracle properties. For the SCAD penalty, although it has the oracle properties, the corresponding optimization problems become nonconvex and thus much harder to solve. Fan and Li [3] proposed the local quadratic

CONTACT L. Y. Fu fuliya@mail.xjtu.edu.cn

approximation (LQA) to solve the optimization problem. Hunter and Li [5] solved the optimization problem using the minimization–maximization (MM) algorithm and analysed the relationship between the MM algorithm and the LQA.

The above studies are all based on the linear regression models. Koenker and Bassett [6] proposed the quantile regression (QR) model, which is more efficient than mean regression when the errors follow a heavy-tailed distribution. QR is robust to outliers in response variable and is capable of dealing with heteroscedasticity, the situation when variances depend on some covariates. More importantly, QR can give a more complete picture on how the responses are affected by covariates. Therefore, QR is widely used to obtain robust estimators and analyse the data from various fields. Then, the above variable selection methods for linear regression models were further extended to QR models. Wu and Liu [7] studied the penalized QR with SCAD and adaptive-Lasso penalties and solved the optimization problems using a difference convex algorithm. Zou and Li [8] also transformed the penalized quantile objective function into a linear programming problem to obtain the estimators based on the local linear approximation. Peng and Wang [9] proposed a new iterative coordinate descent method to solve the optimization problem, which significantly improves the calculation speed when the parameter dimension is large. Fan et al. [10] proposed the adaptive robust lasso (AR-Lasso) method based on QR and demonstrated its excellent variable selection effects. The studies mentioned above all deal with the independent data. Kato [11] studied estimation in functional linear QR in which the dependent variable is scalar while the covariate is a function and also fully observed. Ma et al. [12] studied the high-dimensional functional partially linear quantile model with multiple functional predictors and ultrahigh-dimensional scalar covariates, and developed a double penalized quantile objective function with two nonconvex penalties to select important covariates. Zhang et al. [13] investigated a functional additive QR that models the conditional quantile of a scalar response by nonparametric effects of a functional predictor, and selected the additive components by the SCAD. In the functional QR, the response from the same sample is scalar, and the covariate is a function. The data are independent identically distributed.

Longitudinal data track the same subject at different points in time and are quite common in research experiments like social-personality and clinical psychology. It should be noted that the observations from the same subject are a vector and correlated in the longitudinal data [14], while the data from different subjects are independent, which are different from the functional data. For the variable selection problem with the longitudinal data, Fan and Li [15] and Wang et al. [16] studied the semiparametric and nonparametric marginal models with continuous response variables under the regularization framework. Ni et al. [17] studied the variable selection problem of mixed-effect models with continuous response variables. Xu et al. [18] proposed a variable selection and parameter estimation method using a quasi-objective function based on the generalized estimation equations (GEEs). Johnson et al. [19] showed the asymptotic theory of the penalized GEE. Wang et al. [20] derived the asymptotic theory of the penalized GEE when the dimension of covariates is diverging.

For QR with longitudinal data, Koenker [21] studied the mixed-effect QR models to obtain the sparse random effects. Wang and Fygenson [22] developed inference procedures for longitudinal data where some of the measurements are censored by fixed constants. Leng and Zhang [23] proposed the smoothing combined estimating equations for parameter estimation. Zhang et al. [24] proposed a new QR-based clustering method for panel data. These studies did not consider variable selection. The work of variable selection in the QR with longitudinal data is quite limited due to the complexity of the correlated measurements. Based on the AR Lasso proposed by Fan et al. [10], Gao and Liu [25] proposed the weighted adaptive robust Lasso (WAR-Lasso) to process the ultra-high-dimensional longitudinal data. However, the WAR-Lasso still takes the form of penalized QR and thus ignores the correlations in the longitudinal data. Fu and Wang [26], Wang [27] and Zhu et al. [28] all indicated that assuming the independence working model in the longitudinal data could lead to inefficient estimation.

In this paper, we aim to develop an efficient estimation and variable selection method to deal with the longitudinal data in quantile loss function settings. Instead of using the penalized QR framework, we construct the quantile estimating equations (QEEs) based on the idea of GEEs. The QEE can easily incorporate the correlation structure in the longitudinal data and thus improve the efficiency of parameter estimation and variable selection. To further promote the robustness of the method, we consider integrating the weights of covariates to reduce the impacts of high leverage points. Then the SCAD penalty is added to the weighted quantile estimating equations (WQEEs) to encourage shrinkage and give sparse estimators. For the problem of non-differentiability of the penalized weighted quantile estimating equation (PWQEE), we adopt the local distribution approximation to smooth the functions and solve the smoothed PWQEE using the Newton–Raphson algorithm combined with the MM algorithm.

The rest of this paper is organized as follows. Section 2 introduces the penalized QEEs and the algorithms for handling the corresponding equations. Section 3 presents the numerical results of the simulation. Section 4 illustrates the proposed method by analysing the yeast cell cycle gene expression data. Section 5 gives some conclusions.

2. Methodology and main results

2.1. PWQEEs

Suppose that the longitudinal data are collected from N subjects, and the ith subject has n_i observations for $i = 1, \dots, N$. Let $(y_{ij}, \boldsymbol{x_{ij}})$ denote the jth observation of the ith subject. At a given quantile level $\tau \in (0, 1)$, we consider the QR model as follows:

$$y_{ij} = \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}} \boldsymbol{\beta}(\tau) + \epsilon_{ij}(\tau), \quad i = 1, \dots, N, \ j = 1, \dots, n_i, \tag{1}$$

where $\boldsymbol{x_{ij}}$ is the d-dimensional covariate vector, $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), \dots, \beta_d(\tau))^{\mathrm{T}}$ is the corresponding d-dimensional vector of regression coefficients and $\epsilon_{ij}(\tau)$ is the random error which satisfies $\mathbb{P}(\epsilon_{ij}(\tau) \le 0 | \boldsymbol{x_{ij}}) = \tau$. Assume that ϵ_{ij} are correlated within the same subject and independent between subjects. The interest of this paper is to find a sparse, robust and efficient estimation of $\boldsymbol{\beta}$ for a particular τ. For simplicity, we omit τ in $\boldsymbol{\beta}(\tau)$ and $\epsilon_{ij}(\tau)$, but it should be noted that the quantile level is τ-fixed.

A simple approach to obtain an estimate of $\boldsymbol{\beta}$ is to minimize the following quantile objective function:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left\{ \rho_\tau \left(y_{ij} - \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}} \boldsymbol{\beta} \right) \right\}, \tag{2}$$

where $\rho_\tau(t) = t(\tau - I(t \le 0))$ is the check function. The estimating equations derived from (2) are

$$\boldsymbol{W}_0(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \boldsymbol{x}_{\boldsymbol{ij}} S_{ij} = \sum_{i=1}^{N} \boldsymbol{X}_{\boldsymbol{i}}^{\mathrm{T}} \boldsymbol{S}_{\boldsymbol{i}} = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{X}_{\boldsymbol{i}}$ is the $n_i \times d$ design matrix of the ith subject, whose jth row is $\boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}$, $\boldsymbol{S}_{\boldsymbol{i}} = (S_{i1}, \dots, S_{in_i})^{\mathrm{T}}$ and $S_{ij} = \tau - I(y_{ij} - \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} \le 0)$, where $I(\cdot)$ is the indicator function.

The estimating equations (3) are based on the independence working structure assumption [29]. Although consistent estimators can be obtained by solving (3), they may be inefficient when the errors are correlated. In addition, the estimators gained from (3) are robust to the outliers in the response variable, but sensitive to the high leverage points in the covariates. Therefore, to improve the efficiency and robustness of the parameter estimator, we construct the following WQEEs by incorporating the within correlations and adding weights to the covariates:

$$\boldsymbol{W}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{X}_{\boldsymbol{i}}^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{i}} \boldsymbol{V}_{\boldsymbol{i}}^{-1} \boldsymbol{\Omega}_{\boldsymbol{i}} \boldsymbol{S}_{\boldsymbol{i}} = \boldsymbol{0}, \tag{4}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{i}} = \mathrm{diag}(f_{i1}(0), \dots, f_{in_i}(0))$, in which $f_{ij}(\cdot)$ is the probability density function of ϵ_{ij}, $\boldsymbol{V}_{\boldsymbol{i}}$ is the covariance matrix of $\boldsymbol{S}_{\boldsymbol{i}}$ and can be decomposed to $\boldsymbol{A}_{\boldsymbol{i}}{}^{1/2}\boldsymbol{R}_{\boldsymbol{i}}\boldsymbol{A}_{\boldsymbol{i}}{}^{1/2}$, where $\boldsymbol{A}_{\boldsymbol{i}} = \tau(1-\tau)\boldsymbol{I}_{\boldsymbol{n_i}}$ and $\boldsymbol{R}_{\boldsymbol{i}}$ is the true correlation matrix of $\boldsymbol{S}_{\boldsymbol{i}} = (\tau - I(\epsilon_{i1} \le 0), \dots, \tau - I(\epsilon_{in_i} \le 0))^{\mathrm{T}}$. The matrix $\boldsymbol{\Omega}_{\boldsymbol{i}} = \mathrm{diag}(\omega_{i1}, \dots, \omega_{in_i})$ is a weight function to reduce the influence of high leverage points. In this way, the proposed method is robust to the outliers in both responses and covariates. We set the weights as follows [30]:

$$\omega_{ij}(\boldsymbol{x}_{\boldsymbol{ij}}) = \min\left\{1, \left[\frac{c}{d_{ij}^2(\boldsymbol{x}_{\boldsymbol{ij}})}\right]^{k/2}\right\}, \tag{5}$$

where c and k are tuning constants and $d_{ij}(\boldsymbol{x}_{\boldsymbol{ij}})$ denotes the Mahalanobis distance of $\boldsymbol{x}_{\boldsymbol{ij}}$ based on some robust measure of location and dispersion for the total design set $\boldsymbol{X}$. Here, we calculate $d_{ij}(\boldsymbol{x}_{\boldsymbol{ij}})$ using the fast minimum covariance determinant estimates proposed by Rousseeuw and Driessen [31]. These estimates are available through the **MASS** package in the statistical software R. For the tuning constants, we set $c = \chi^2_{0.95}(d)$, the 95th percentile of a $\chi^2(d)$ distribution, where d is the dimensions of $\boldsymbol{\beta}$; and $k = 2$ as Terpstra and McKean [30] suggested. If random errors are identically distributed, $\boldsymbol{\Lambda}_{\boldsymbol{i}}$ becomes a constant matrix and thus can be ignored.

According to Fu and Wang [32], we construct the working correlation matrix $\boldsymbol{R}_{\boldsymbol{i}} = (\gamma_{i,jk})$ via Gaussian copulas, where $\gamma_{i,jk}$ is the correlation coefficient between S_{ij} and S_{ik}. Suppose that $F_{i,jk}(\cdot,\cdot)$ is the joint cumulative distribution function of $(\epsilon_{ij}, \epsilon_{ik})$ with marginal distributions $F_{ij}(\cdot)$ and $F_{ik}(\cdot)$. Because $\mathrm{Var}(S_{ij}) = \tau(1-\tau)$, we have

$$\begin{aligned}\gamma_{i,jk} &= \frac{1}{\tau(1-\tau)}\left[\mathbb{P}\left(\epsilon_{ij} \le 0, \epsilon_{ik} \le 0\right) - \tau^2\right] \\ &= \frac{1}{\tau(1-\tau)}\left[\mathbb{P}\left(F_{ij}(\epsilon_{ij}) \le F_{ij}(0), F_{ik}(\epsilon_{ik}) \le F_{ik}(0)\right) - \tau^2\right]\end{aligned}$$

$$= \frac{1}{\tau(1-\tau)}\left[C_{i,jk}(F_{ij}(0), F_{ik}(0)) - \tau^2\right]$$
$$= \frac{C_{i,jk}(\tau,\tau) - \tau^2}{\tau(1-\tau)}, \tag{6}$$

where $C_{i,jk}(\tau,\tau) = \Phi_2(\Phi^{-1}(\tau), \Phi^{-1}(\tau), \rho_{i,jk})$ is a Gaussian copula and $\Phi_2(\cdot,\cdot,\rho_{i,jk})$ is the standardized bivariate normal distribution with correlation coefficient $\rho_{i,jk}$, $\Phi^{-1}(\cdot)$ is the inverse function of the univariate standardized normal distribution and $\rho_{i,jk}$ denotes the correlation coefficient between ϵ_{ij} and ϵ_{ik} and can be estimated by the moment method [33]. In this paper, we consider three kinds of working correlation structures:

(i) independence correlation matrix $\boldsymbol{R}_{\boldsymbol{i},\mathbf{id}} = I_{n_i}$;
(ii) exchangeable correlation matrix $\boldsymbol{R}_{\boldsymbol{i},\mathbf{ex}}$, where $C_{i,jk}(\tau,\tau) = \Phi_2(\Phi^{-1}(\tau), \Phi^{-1}(\tau), \rho)$;
(iii) AR(1) correlation matrix $\boldsymbol{R}_{\boldsymbol{i},\mathbf{ar}}$, where $C_{i,jk}(\tau,\tau) = \Phi_2(\Phi^{-1}(\tau), \Phi^{-1}(\tau), \rho^{|j-k|})$.

Suppose that the true parameter is sparse, we utilize the regularization to identify the important covariates and propose the PWQEEs:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{W}(\boldsymbol{\beta}) - N\boldsymbol{q}_\lambda(|\boldsymbol{\beta}|) \circ \mathrm{sgn}(\boldsymbol{\beta}) = \mathbf{0}, \tag{7}$$

where $\boldsymbol{q}_\lambda(|\boldsymbol{\beta}|) = (q_\lambda(|\beta_1|), \ldots, q_\lambda(|\beta_d|))^{\mathrm{T}}$, $\mathrm{sgn}(\boldsymbol{\beta}) = (\mathrm{sgn}(\beta_1), \ldots, \mathrm{sgn}(\beta_d))^{\mathrm{T}}$ and $\mathrm{sgn}(x) = I(x>0) - I(x<0)$. The notation $\boldsymbol{q}_\lambda(|\boldsymbol{\beta}|) \circ \mathrm{sgn}(\boldsymbol{\beta})$ denotes the component-wise product. The tuning parameter λ determines the amount of shrinkage. The penalized function adopts the SCAD penalty given by

$$q_\lambda(|\beta|) = p'_\lambda(|\beta|) = \lambda I(|\beta| \le \lambda) + \frac{(a\lambda - |\beta|)_+}{a-1} I(|\beta| > \lambda), \quad a > 2.$$

Note that the penalized weighted quantile estimating function is noncontinuous. The discontinuity brought by the indicator function $I(y_{ij} - \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} \le 0)$ in $\boldsymbol{S_i}$ poses a challenge in solving Equation (7). To overcome this obstacle, we adopt a local distribution function proposed by Heller [34] to approximate the indicator function and thus smooth the penalized weighted quantile estimating function. The local distribution function is defined by $\Phi((\boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} - y_{ij})/h)$, where h is a bandwidth parameter. Note that if $y_{ij} - \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} \le 0$, $\Phi((\boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} - y_{ij})/h) \to 1$ as h tends to zero; while $\Phi((\boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} - y_{ij})/h) \to 0$ if $y_{ij} - \boldsymbol{x}_{\boldsymbol{ij}}^{\mathrm{T}}\boldsymbol{\beta} \ge 0$. Then, the smoothed PWQEE become

$$\widetilde{\boldsymbol{U}}(\boldsymbol{\beta}) = \widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - N\boldsymbol{q}_\lambda(|\boldsymbol{\beta}|) \circ \mathrm{sgn}(\boldsymbol{\beta})$$
$$= \sum_{i=1}^{N} \boldsymbol{X}_{\boldsymbol{i}}^{\mathrm{T}} \boldsymbol{V}_{\boldsymbol{i}}^{-1} \boldsymbol{\Omega}_{\boldsymbol{i}} (\boldsymbol{\tau} - \boldsymbol{\Phi}_{\boldsymbol{i}}(\boldsymbol{\beta})) - N\boldsymbol{q}_\lambda(|\boldsymbol{\beta}|) \circ \mathrm{sgn}(\boldsymbol{\beta}) = \mathbf{0}, \tag{8}$$

where $\boldsymbol{\tau} - \boldsymbol{\Phi}_{\boldsymbol{i}}(\boldsymbol{\beta}) = (\tau - \Phi((\boldsymbol{x}_{\boldsymbol{i1}}^{\mathrm{T}}\boldsymbol{\beta} - y_{i1})/h), \ldots, \tau - \Phi((\boldsymbol{x}_{\boldsymbol{in_i}}^{\mathrm{T}}\boldsymbol{\beta} - y_{in_i})/h))^{\mathrm{T}}$ is a n_i-dimensional vector. Here we take $h = \hat{\sigma} M^{-0.26}$ as Heller [34] suggested, where $\hat{\sigma}$ is the standard deviation of the residuals, and $M = \sum_{i=1}^{N} n_i$ is the total number of measurements.

A simulation is conducted to show the smooth effects of the local distribution function (see Figure 1). The data are generated from $y_i = \boldsymbol{x}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_i$, where the covariates $\boldsymbol{x_i} \sim$

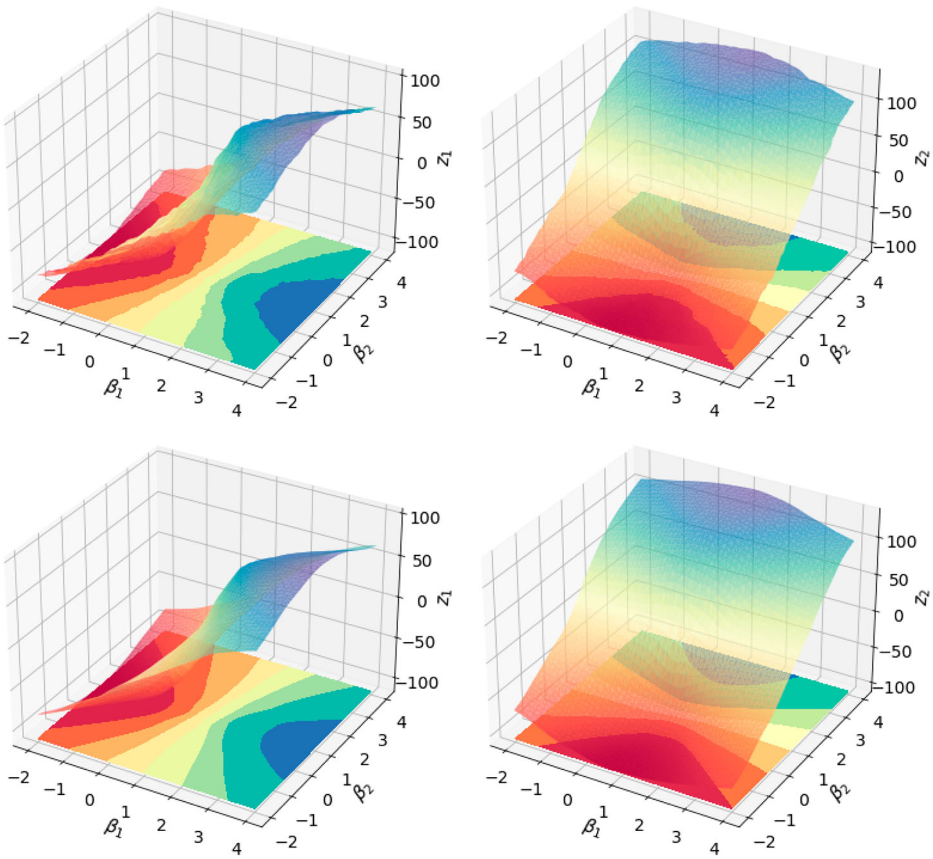

Figure 1. A simulation about the local distribution function.

$N_2(\mathbf{0}, \boldsymbol{I}_2)$, the regression coefficients $\boldsymbol{\beta} = (1,1)^{\mathrm{T}}$ and the random error $\epsilon_i \sim N(0,1)$, $\quad i = 1,\ldots,300$. The two subplots above in Figure 1 show the two components of $\boldsymbol{X}^{\mathrm{T}}(I(y_1 - \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\beta} \le 0),\ldots,I(y_{300} - \boldsymbol{x}_{300}^{\mathrm{T}}\boldsymbol{\beta} \le 0))^{\mathrm{T}}$, respectively, and the two subplots below in Figure 1 show the two components of $\boldsymbol{X}^{\mathrm{T}}(\Phi((\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\beta} - y_1)/h),\ldots,\Phi((\boldsymbol{x}_{300}^{\mathrm{T}}\boldsymbol{\beta} - y_{300})/h))^{\mathrm{T}}$, where $h = \hat{\sigma} \times 300^{-0.26}$.

The smoothed penalized weighted quantile estimating functions are nonconvex. Similarly as in Ref. [19], we combine the Newton–Raphson algorithm and the MM algorithm for the nonconvex penalty of Hunter and Li [5] to overcome this difficulty. For a small $\eta > 0$, the MM algorithm gives that the estimated regression coefficients $\widetilde{\boldsymbol{\beta}} = (\tilde{\beta}_1,\ldots,\tilde{\beta}_d)^{\mathrm{T}}$ approximately satisfy

$$\widetilde{W}_j(\widetilde{\boldsymbol{\beta}}) \approx Nq_\lambda(|\tilde{\beta}_j|)\mathrm{sgn}(\tilde{\beta}_j)\frac{|\tilde{\beta}_j|}{\eta + |\tilde{\beta}_j|}, \quad j = 1,\ldots,d, \tag{9}$$

where $\widetilde{W}_j(\widetilde{\boldsymbol{\beta}})$ is the jth component of $\widetilde{\boldsymbol{W}}(\widetilde{\boldsymbol{\beta}})$. In practice, we take η to be a fixed small value 10^{-6}. Applying the Newton–Raphson algorithm to $\widetilde{\boldsymbol{W}}(\boldsymbol{\beta})$ and combining with (9), we obtain the following iterative algorithm to solve the PWQEE, see Algorithm 1.

Algorithm 1 (The Newton–Raphson algorithm combining with MM)

Step 0: Calculate the weights $\omega_{ij}, i = 1, \ldots, N, j = 1, \ldots, n_i$. Initialize $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}_\tau$, where $\tilde{\boldsymbol{\beta}}_\tau$ is an estimator obtained by the weighted QR, which is available in R through **rq** function in **quantreg** package. Initialize $\sigma^{(0)} = \mathrm{SD}\left(\boldsymbol{X}\boldsymbol{\beta}^{(0)} - \boldsymbol{y}\right)$, thus $h^{(0)} = \sigma^{(0)} M^{-0.26}$. Estimate $\rho^{(0)}$ through the moment method and initialize $\boldsymbol{R_i}(\boldsymbol{\beta}^{(0)})$.
Step 1: For a given $\boldsymbol{\beta}^{(t)}, \boldsymbol{R_i}(\boldsymbol{\beta}^{(t)})$ and $h^{(t)}, t = 0, 1, \ldots$, update $\boldsymbol{\beta}^{(t+1)}$ by

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left[\boldsymbol{H}(\boldsymbol{\beta}^{(t)}) + N\boldsymbol{E}(\boldsymbol{\beta}^{(t)})\right]^{-1} \times \left[\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}^{(t)}) - N\boldsymbol{E}(\boldsymbol{\beta}^{(t)})\boldsymbol{\beta}^{(t)}\right], \tag{10}$$

where

$$\boldsymbol{H}(\boldsymbol{\beta}^{(t)}) = \sum_{i=1}^{N} \frac{1}{h^{(t)}} \boldsymbol{X_i}^{\mathrm{T}} \boldsymbol{V_i}^{-1}(\boldsymbol{\beta}^{(t)}) \boldsymbol{\phi}(\boldsymbol{\beta}^{(t)}) \boldsymbol{\Omega}_i \boldsymbol{X_i}, \tag{11}$$

$$\boldsymbol{V_i}^{-1}(\boldsymbol{\beta}^{(t)}) = \boldsymbol{A_i}^{-1/2} \boldsymbol{R_i}^{-1}(\boldsymbol{\beta}^{(t)}) \boldsymbol{A_i}^{-1/2}, \tag{12}$$

$$\boldsymbol{\phi}(\boldsymbol{\beta}^{(t)}) = \mathrm{diag}\left\{\phi\left(\frac{\boldsymbol{x}_{i1}^{\mathrm{T}}\boldsymbol{\beta}^{(t)} - y_{i1}}{h^{(t)}}\right), \ldots, \phi\left(\frac{\boldsymbol{x}_{in_i}^{\mathrm{T}}\boldsymbol{\beta}^{(t)} - y_{in_i}}{h^{(t)}}\right)\right\}, \tag{13}$$

$$\boldsymbol{E}(\boldsymbol{\beta}^{(t)}) = \mathrm{diag}\left\{\frac{q_\lambda(|\beta_1^{(t)}|)}{\eta + |\beta_1^{(t)}|}, \ldots, \frac{q_\lambda(|\beta_d^{(t)}|)}{\eta + |\beta_d^{(t)}|}\right\}, \tag{14}$$

where $\phi(x)$ is the probability density function of the standard normal distribution.
Step 2: For a given $\boldsymbol{\beta}^{(t+1)}, t = 0, 1, \ldots$, update $h^{(t+1)}$, $\rho^{(t+1)}$ and $\boldsymbol{R_i}(\boldsymbol{\beta}^{(t+1)})$ through the way in Step 0.
Step 3: Repeat Step 1 and Step 2 until convergence. If $\sum_{i=1}^{d} |\beta_i^{(t+1)} - \beta_i^{(t)}| < 10^{-3}$, then stop the iteration.

2.2. Asymptotic properties

In this subsection, we study the asymptotic properties of the proposed method. Let $\boldsymbol{\beta_0} = (\beta_{01}, \ldots, \beta_{0d})^{\mathrm{T}}$ denote the true value of $\boldsymbol{\beta}$. Without loss of generality, suppose that $\beta_{0j}, j = 1, \ldots, s$, are the non-zero components of $\boldsymbol{\beta_0}$ and $\beta_{0j} = 0$ for $j = s+1, \ldots, d$. To derive the oracle property of PWQEE, we need the following conditions and lemmas.

(A1) The cumulative distribution functions $F_{ij}(\cdot)$ of ϵ_{ij} are absolutely continuous. The corresponding density function $f_{ij}(\cdot)$ is continuous and uniformly bounded, and $f'_{ij}(\cdot)$ exists and is uniformly bounded.
(A2) The true coefficients $\boldsymbol{\beta_0}$ are in the interior of a bounded convex region $\mathbb{B}$.
(A3) For any positive definite matrix $\boldsymbol{C_i}$, $N^{-1}\sum_{i=1}^{N} \boldsymbol{X_i}^{\mathrm{T}} \boldsymbol{C_i} \boldsymbol{\Omega_i} \boldsymbol{\Lambda_i} \boldsymbol{X_i}$ converges to a positive definite matrix, where $\boldsymbol{\Lambda_i} = \mathrm{diag}(f_{i1}(0), \ldots, f_{in_i}(0))$ and $\sup_i ||\boldsymbol{X_i}|| < +\infty$.
(A4) The bandwidth h is chosen such as $N \to \infty$, $Nh \to \infty$ and $Nh^4 \to 0$.
(A5) The tuning parameter $\lambda \to 0$ and $\sqrt{N}\lambda \to \infty$.

A1 is a standard assumption about the distribution in QR. A2 guarantees the existence of parameter estimates. A3 ensures that the covariance matrix of the estimator is positive definite. A4 and A5 require selecting appropriate parameters h and λ to ensure asymptotic properties.

Lemma 2.1: *Let* $\boldsymbol{A} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i\boldsymbol{X}_i$ *and* $\Xi = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{\Omega}_i\boldsymbol{V}_i\boldsymbol{\Omega}_i\boldsymbol{V}_i^{-1}\boldsymbol{X}_i$. *Under conditions A1–A3, for a given constant* $K > 0$*, we have* $N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta}_0) \to N(\mathbf{0}, \Xi)$ *and*

$$\sup_{||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\le KN^{-1/2}} ||N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta}) - N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta}_0) - N^{1/2}\boldsymbol{A}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)|| = o_p(1). \tag{12}$$

Lemma 2.2: *Under conditions A1–A4,* $N^{-1/2}\{\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta})\} = o_p(1)$.

Lemma 2.1 shows the asymptotic properties of the weighted quantile estimating functions. Furthermore, the normality of the corresponding estimator obtained from $\boldsymbol{W}(\boldsymbol{\beta}) = \mathbf{0}$ can be easily derived from (12). Lemma 2.2 indicates that the unsmoothed and smoothed estimating functions are asymptotically equivalent in $\boldsymbol{\beta}$, which indicates that the smoothing method hardly changes the estimating equations except making them differentiable. Proofs of the lemmas and the following theorem are provided in the Appendix.

Theorem 2.3: *Denote the number of non-zero components* $s = \#\{j|\beta_{0j} \neq 0\}$. *Under conditions A1–A5, the following results hold:*

(a) *The smoothed estimator,* $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(N^{-1/2})$.
(b) *For the root-N-consistent solution* $\widetilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \dots, \tilde{\beta}_d)^{\mathrm{T}}$ *to* $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta})$. *Let* $\widetilde{\boldsymbol{\beta}}_1 = (\tilde{\beta}_1, \dots, \tilde{\beta}_s)^{\mathrm{T}}$, $\boldsymbol{\beta}_{01} = (\beta_{01}, \dots, \beta_{0s})^{\mathrm{T}}$ *and* $\widetilde{\boldsymbol{\beta}}_2 = (\tilde{\beta}_{s+1}, \dots, \tilde{\beta}_d)^{\mathrm{T}}$. *We have*

$$\lim_{N\to\infty} \mathbb{P}(\widetilde{\boldsymbol{\beta}}_2 = \mathbf{0_{d-s}}) = \mathbf{1}$$

and

$$N^{1/2}\left(\boldsymbol{A}_{11} + \boldsymbol{\Sigma}_{11}\right)\left\{\widetilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{01}\right\} \to N\left(\mathbf{0}, \Xi_{11}\right), \tag{13}$$

where $\boldsymbol{A}_{11}$, $\boldsymbol{\Sigma}_{11}$ *and* Ξ_{11} *are the first* $s \times s$ *submatrices of* $\boldsymbol{A}$, $\mathrm{diag}\{-q'_\lambda(|\boldsymbol{\beta}_0|)\mathrm{sgn}(\boldsymbol{\beta}_0)\}$ *and* Ξ*, respectively.*

This theorem states the main asymptotic properties of the smoothed PWQEE estimators, including the existence of a root-N-consistent estimator, the sparsity and asymptotic normality of the estimators.

3. Simulation studies

In this section, we conduct numerical simulations to evaluate the performance of the model selection and parameter estimation of the PWQEE. An estimate is treated as zero if its absolute value is less than 10^{-4}. For the SCAD penalty, we set $a = 3.7$ as in Ref. [3] to reduce the computational burden. The tuning parameter λ is chosen as the minimizer of this following BIC (Bayesian information criterion)-type criterion [35]:

$$\mathrm{BIC}_\lambda = \log\left\{\frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\rho_\tau(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\lambda)\right\} + DF_\lambda\frac{\log M}{M}, \tag{14}$$

where DF_λ denotes the number of non-zero components in $\widehat{\boldsymbol{\beta}}_\lambda$.

The data are generated from the following model:

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon_{ij}, \quad i = 1,\ldots,200, \;\; j = 1,\ldots,4, \tag{15}$$

where $\boldsymbol{\beta}_0 = (3, 1.5, 2, 0_{p-3})^{\mathrm{T}}$, the components of $\boldsymbol{x}_{ij}$ follow the standard normal distribution. The correlation between any two components x_{ijk} and x_{ijh} is set to $\rho^{|k-h|}$ with $\rho = 0.5$. The error term $\epsilon_{ij} = \xi + e_{ij}$, where ξ is a shift to ensure $\mathbb{P}(\epsilon_{ij} \le 0) = \tau$. For the dimension of β, we consider $p = 8$ and $p = 100$. Three cases are considered for e_{ij} and $\boldsymbol{x}_{ij}$:

Case 1: Multivariate normal distribution. Assume that $\boldsymbol{e}_i = (e_{i1},\ldots,e_{in_i})^{\mathrm{T}} \sim$ MVN $(\boldsymbol{0}, \boldsymbol{\Sigma})$, where the correlation structure, $\boldsymbol{\Sigma}$, is AR(1) with correlation parameter α, i.e. $\Sigma_{jj'} = \alpha^{|j-j'|}$.

Case 2: Multivariate t distribution. We assume that $\boldsymbol{e}_i = (e_{i1},\ldots,e_{in_i})^{\mathrm{T}}$ follows a multivariate t-distribution with three degrees of freedom, $\mathrm{MVT}_3(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an AR(1) matrix with correlation parameter α, i.e. $\Sigma_{jj'} = \alpha^{|j-j'|}$.

Table 1. Variable selection results for case 1.

		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
		T	R	W	E	T	R	W	E	T	R	W	E
$p = 8$													
	PWQEE.id	0.318	4.77	0	0.86	0.400	4.84	0	0.88	0.319	4.78	0	0.83
	PWQEE.ex	0.317	4.95	0	0.98	0.399	5	0	1	0.318	4.97	0	0.98
$\alpha = 0.9$	PWQEE.ar	0.317	4.96	0	0.98	0.399	4.97	0	0.98	0.317	4.99	0	0.99
	PQEE.ar	0.317	4.96	0	0.98	0.399	4.98	0	0.99	0.317	4.98	0	0.99
	oracle	0.317	5	0	1	0.399	5	0	1	0.317	5	0	1
	PWQEE.id	0.318	4.8	0	0.85	0.400	4.84	0	0.86	0.319	4.72	0	0.78
	PWQEE.ex	0.317	4.93	0	0.96	0.400	4.97	0	0.97	0.319	4.91	0	0.93
$\alpha = 0.5$	PWQEE.ar	0.317	4.88	0	0.90	0.400	4.88	0	0.9	0.319	4.82	0	0.86
	PQEE.ar	0.317	4.89	0	0.93	0.400	4.89	0	0.91	0.319	4.79	0	0.84
	oracle	0.317	5	0	1	0.400	5	0	1	0.318	5	0	1
	PWQEE.id	0.317	4.82	0	0.87	0.402	4.88	0	0.91	0.319	4.74	0	0.82
	PWQEE.ex	0.317	4.94	0	0.97	0.402	4.97	0	0.97	0.319	4.9	0	0.94
$\alpha = 0$	PWQEE.ar	0.317	4.82	0	0.87	0.402	4.88	0	0.91	0.319	4.74	0	0.82
	PQEE.ar	0.317	4.85	0	0.89	0.402	4.89	0	0.91	0.319	4.74	0	0.83
	oracle	0.317	5	0	1	0.402	5	0	1	0.319	5	0	1
$p = 100$													
	PWQEE.id	0.322	95.41	0	0.70	0.399	96.53	0	0.88	0.319	95.57	0	0.74
	PWQEE.ex	0.320	96.88	0	0.88	0.398	96.97	0	0.97	0.317	96.92	0	0.92
$\alpha = 0.9$	PWQEE.ar	0.319	97	0	1	0.398	96.99	0	0.99	0.317	97	0	1
	PQEE.ar	0.319	97	0	1	0.398	97	0	1	0.316	97	0	1
	oracle	0.319	97	0	1	0.398	97	0	1	0.316	97	0	1
	PWQEE.id	0.322	95.57	0	0.75	0.399	96.75	0	0.92	0.322	95.50	0	0.68
	PWQEE.ex	0.319	96.96	0	0.96	0.399	96.98	0	0.98	0.319	96.90	0	0.94
$\alpha = 0.5$	PWQEE.ar	0.320	96.08	0	0.90	0.399	96.93	0	0.98	0.320	96.22	0	0.84
	PQEE.ar	0.320	96.34	0	0.91	0.399	96.90	0	0.97	0.320	96.37	0	0.86
	oracle	0.319	97	0	1	0.399	97	0	1	0.319	97	0	1
	PWQEE.id	0.320	95.38	0	0.70	0.400	96.59	0	0.88	0.322	95.50	0	0.68
	PWQEE.ex	0.318	96.98	0	0.98	0.399	96.98	0	0.99	0.320	96.98	0	0.98
$\alpha = 0$	PWQEE.ar	0.320	95.35	0	0.70	0.400	96.62	0	0.90	0.321	95.92	0	0.82
	PQEE.ar	0.320	95.30	0	0.74	0.400	96.73	0	0.91	0.321	96.35	0	0.84
	oracle	0.318	97	0	1	0.399	97	0	1	0.320	97	0	1

Case 3: Contaminated covariates. Assume that $\boldsymbol{e_i} = (e_{i1}, \ldots, e_{in_i})^{\mathrm{T}} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\Sigma_{jj'} = \alpha^{|j-j'|}$. The covariates of 5% of the subjects are randomly contaminated by random numbers generated from an uniform distribution $U[5, 10]$.

We evaluate the performance of the proposed PWQEE method with different correlation structures and verify the improvement of robustness by adding weights to covariates. Here, five methods are considered: PWQEE method with the independence/exchangeable/AR(1) correlation structure (abbreviated as PWQEE.id/ex/ar), PQEE method with the AR(1) correlation structure (abbreviated as PQEE.ar, which can be obtained by setting all $\omega_{ij} = 1$) and the WQEE.ar-oracle method (abbreviated as oracle).

For each case, we compare the performance of variable selection and parameter estimation of five methods under $\alpha = 0.9, 0.5$ and 0 and different quantile levels ($\tau = 0.25, 0.5, 0.75$). In the same way as the training sets, we also generate testing sets to test the forecasting ability. After 100 replicates, the variable selection results of the five methods are shown in Tables 1–3, with the top panels for $p = 8$ and the bottom panels for $p = 100$, respectively. The reported test error (T) refers to the average check loss on the independent testing data set. The notation 'R' denotes the average number of zero regression coefficients

Table 2. Variable selection results for case 2.

		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
		T	R	W	E	T	R	W	E	T	R	W	E
$p = 8$													
	PWQEE.id	0.469	4.83	0	0.86	0.553	4.95	0	0.96	0.461	4.85	0	0.86
	PWQEE.ex	0.468	4.97	0	0.97	0.552	5	0	1	0.460	5	0	1
$\alpha = 0.9$	PWQEE.ar	0.468	4.98	0	0.98	0.552	5	0	1	0.460	5	0	1
	PQEE.ar	0.468	4.99	0	0.99	0.552	5	0	1	0.460	5	0	1
	oracle	0.468	5	0	1	0.552	5	0	1	0.460	5	0	1
	PWQEE.id	0.461	4.88	0	0.89	0.554	4.94	0	0.94	0.462	4.83	0	0.84
	PWQEE.ex	0.460	4.92	0	0.95	0.554	4.99	0	0.99	0.461	4.99	0	0.99
$\alpha = 0.5$	PWQEE.ar	0.460	4.87	0	0.90	0.554	4.97	0	0.97	0.462	4.85	0	0.86
	PQEE.ar	0.460	4.88	0	0.91	0.554	4.97	0	0.97	0.461	4.89	0	0.89
	oracle	0.460	5	0	1	0.554	5	0	1	0.461	5	0	1
	PWQEE.id	0.465	4.77	0	0.82	0.553	4.95	0	0.95	0.462	4.82	0	0.84
	PWQEE.ex	0.465	4.91	0	0.93	0.553	4.99	0	0.99	0.461	4.96	0	0.96
$\alpha = 0$	PWQEE.ar	0.465	4.76	0	0.81	0.553	4.95	0	0.95	0.462	4.82	0	0.84
	PQEE.ar	0.465	4.76	0	0.79	0.553	4.94	0	0.94	0.462	4.82	0	0.84
	oracle	0.464	5	0	1	0.553	5	0	1	0.461	5	0	1
$p = 100$													
	PWQEE.id	0.462	95.73	0	0.61	0.549	96.9	0	0.94	0.469	95.64	0	0.69
	PWQEE.ex	0.459	96.92	0	0.92	0.549	97	0	1	0.465	96.93	0	0.93
$\alpha = 0.9$	PWQEE.ar	0.458	96.99	0	0.99	0.548	97	0	1	0.465	96.96	0	0.99
	PQEE.ar	0.458	97	0	1	0.548	97	0	1	0.465	96.96	0	0.99
	oracle	0.458	97	0	1	0.548	97	0	1	0.465	97	0	1
	PWQEE.id	0.457	95.66	0	0.64	0.550	96.88	0	0.96	0.463	95.57	0	0.64
	PWQEE.ex	0.454	96.96	0	0.97	0.549	96.99	0	0.99	0.460	96.96	0	0.96
$\alpha = 0.5$	PWQEE.ar	0.455	96.30	0	0.89	0.549	96.94	0	0.99	0.461	96.45	0	0.83
	PQEE.ar	0.455	96.45	0	0.91	0.549	96.95	0	0.99	0.461	96.28	0	0.85
	oracle	0.454	97	0	1	0.549	97	0	1	0.460	97	0	1
	PWQEE.id	0.463	96.25	0	0.74	0.553	96.92	0	0.96	0.462	95.61	0	0.59
	PWQEE.ex	0.462	96.92	0	0.93	0.553	96.96	0	0.97	0.459	96.86	0	0.88
$\alpha = 0$	PWQEE.ar	0.463	96.29	0	0.74	0.553	96.87	0	0.94	0.462	95.66	0	0.66
	PQEE.ar	0.463	96.11	0	0.74	0.553	96.95	0	0.97	0.462	95.61	0	0.72
	oracle	0.462	97	0	1	0.553	97	0	1	0.459	97	0	1

that are correctly estimated as zero (the optimal value is 5); ‘W’ denotes the average number of non-zero regression coefficients that are incorrectly estimated as zero (the optimal value is 0); ‘E’ is the proportion of the replicates where the model is correctly identified.

We can find that when the errors are multivariate normal (case 1, Table 1) or multivariate t (case 2, Table 2), the five methods give similar test errors. When the correlation parameter α is large, the PWQEE method with the exchangeable correlation structure (PWQEE.ex) or AR(1) structure (PWQEE.ar) outperforms that with the independence structure (PWQEE.id). When the data are not contaminated, PWQEE.ar is comparable with PQEE.ar; while the covariates are contaminated (case 3, Table 3), the model selection effect of PQEE.ar decreases, and its test error is significantly higher than those of the methods adding weights. The proportions of selecting the true model of PWQEE.id decrease as the dimension of covariates increases from $p = 8$ to $p = 100$. When $p = 100$ and the covariates are contaminated (the bottom panel of Table 3), PQEE.ar cannot select the true model.

Tables 4–6 and Tables 7–9 report the biases, mean square errors (MSEs) and the average MSE (abbreviated as AM) of the three non-zero coefficient estimates in 100 replicates for $p = 8$ and $p = 100$, respectively. Since the values of bias and MSE are small, we multiply them by 100 for convenient comparison.

Table 3. Variable selection results for case 3.

		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
		T	R	W	E	T	R	W	E	T	R	W	E
$p = 8$													
	PWQEE.id	0.319	4.81	0	0.84	0.400	4.85	0	0.91	0.319	4.76	0	0.84
	PWQEE.ex	0.318	4.91	0	0.91	0.400	4.89	0	0.91	0.319	4.74	0	0.83
$\alpha = 0.9$	PWQEE.ar	0.318	4.97	0	0.97	0.400	5	0	1	0.319	4.89	0	0.93
	PQEE.ar	0.453	4.75	0	0.75	0.619	4.47	0.09	0.42	0.359	4.68	0.06	0.76
	oracle	0.318	5	0	1	0.400	5	0	1	0.319	5	0	1
	PWQEE.id	0.321	4.74	0	0.79	0.399	4.87	0	0.90	0.319	4.68	0	0.79
	PWQEE.ex	0.320	4.89	0	0.93	0.399	4.91	0	0.94	0.318	4.88	0	0.95
$\alpha = 0.5$	PWQEE.ar	0.321	4.81	0	0.85	0.399	4.85	0	0.89	0.318	4.80	0	0.85
	PQEE.ar	0.445	4.80	0	0.81	0.622	4.52	0.10	0.45	0.359	4.69	0.04	0.71
	oracle	0.320	5	0	1	0.399	5	0	1	0.318	5	0	1
	PWQEE.id	0.319	4.83	0	0.88	0.400	4.89	0	0.91	0.318	4.71	0	0.82
	PWQEE.ex	0.319	4.95	0	0.97	0.400	4.96	0	0.98	0.318	4.90	0	0.95
$\alpha = 0$	PWQEE.ar	0.319	4.81	0	0.87	0.400	4.90	0	0.92	0.318	4.72	0	0.82
	PQEE.ar	0.449	4.85	0	0.86	0.632	4.45	0.1	0.39	0.358	4.65	0.06	0.72
	oracle	0.319	5	0	1	0.400	5	0	1	0.318	5	0	1
$p = 100$													
	PWQEE.id	0.320	95.38	0	0.66	0.406	96.20	0	0.80	0.321	95.20	0	0.69
	PWQEE.ex	0.318	95.54	0	0.67	0.402	96.63	0	0.89	0.320	96.10	0	0.81
$\alpha = 0.9$	PWQEE.ar	0.318	97	0	1	0.401	97	0	1	0.319	96.80	0	0.90
	PQEE.ar	0.580	62.89	0	0	0.579	88.23	0.04	0.23	0.359	93.86	0.04	0.60
	oracle	0.318	97	0	1	0.401	97	0	1	0.319	97	0	1
	PWQEE.id	0.321	95.63	0	0.67	0.401	96.76	0	0.92	0.321	95.27	0	0.70
	PWQEE.ex	0.319	96.84	0	0.88	0.400	96.92	0	0.93	0.318	96.15	0	0.81
$\alpha = 0.5$	PWQEE.ar	0.319	96.53	0	0.89	0.400	96.96	0	0.98	0.318	96.80	0	0.94
	PQEE.ar	0.579	62.53	0.04	0	0.579	89.24	0	0.4	0.358	93.56	0	0.58
	oracle	0.319	97	0	1	0.400	97	0	1	0.318	97	0	1
	PWQEE.id	0.322	95.44	0	0.71	0.400	96.84	0	0.95	0.321	95.6	0	0.73
	PWQEE.ex	0.320	96.96	0	0.98	0.399	96.92	0	0.97	0.319	96.95	0	0.96
$\alpha = 0$	PWQEE.ar	0.322	96.29	0	0.90	0.400	96.75	0	0.94	0.321	96.98	0	0.99
	PQEE.ar	0.579	63.31	0	0	0.578	88.44	0	0.35	0.358	94.1	0	0.65
	oracle	0.320	97	0	1	0.399	97	0	1	0.319	97	0	1

Table 4. Parameter estimation results of non-zero coefficients for case 1 when $p = 8$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
		PWQEE.id	0.508	0.345	0.091	0.388	−0.465	0.250	0.328
		PWQEE.ex	0.419	0.143	−1.269	0.281	0.248	0.122	0.182
	$\alpha = 0.9$	PWQEE.ar	−0.190	0.104	0.171	0.159	−0.333	0.095	0.119
		PQEE.ar	−0.095	0.098	0.142	0.141	−0.349	0.098	0.112
		oracle	−0.174	0.103	0.178	0.159	−0.274	0.091	0.118
		PWQEE.id	0.231	0.269	1.161	0.204	−0.460	0.234	0.236
		PWQEE.ex	0.317	0.238	0.748	0.201	−0.569	0.221	0.220
$\tau = 0.25$	$\alpha = 0.5$	PWQEE.ar	0.225	0.240	0.899	0.217	−0.562	0.212	0.223
		PQEE.ar	0.212	0.235	0.974	0.230	−0.644	0.222	0.229
		oracle	0.266	0.246	0.797	0.228	−0.549	0.209	0.228
		PWQEE.id	−0.303	0.339	−0.915	0.381	0.735	0.357	0.359
		PWQEE.ex	−0.331	0.338	−0.937	0.371	0.762	0.358	0.356
	$\alpha = 0$	PWQEE.ar	−0.315	0.339	−0.912	0.378	0.707	0.354	0.357
		PQEE.ar	−0.325	0.334	−0.895	0.372	0.710	0.351	0.352
		oracle	−0.364	0.341	−0.883	0.372	0.745	0.352	0.355
		PWQEE.id	0.725	0.294	−0.391	0.327	0.243	0.230	0.284
		PWQEE.ex	1.139	0.121	−1.899	0.226	0.761	0.123	0.156
	$\alpha = 0.9$	PWQEE.ar	0.017	0.085	0.291	0.090	−0.027	0.087	0.087
		PQEE.ar	0.122	0.078	0.191	0.083	0.034	0.079	0.080
		oracle	0.015	0.084	0.304	0.092	−0.020	0.085	0.087
		PWQEE.id	0.557	0.249	−0.433	0.216	−0.018	0.251	0.239
		PWQEE.ex	0.579	0.213	−0.493	0.203	−0.008	0.204	0.207
$\tau = 0.5$	$\alpha = 0.5$	PWQEE.ar	0.389	0.212	0.017	0.202	−0.084	0.195	0.203
		PQEE.ar	0.320	0.202	0.042	0.196	−0.127	0.195	0.198
		oracle	0.408	0.212	−0.024	0.203	−0.135	0.192	0.202
		PWQEE.id	0.161	0.198	0.328	0.306	0.453	0.303	0.269
		PWQEE.ex	0.127	0.200	0.331	0.307	0.511	0.287	0.265
	$\alpha = 0$	PWQEE.ar	0.176	0.199	0.331	0.305	0.462	0.306	0.270
		PQEE.ar	0.099	0.194	0.386	0.305	0.485	0.300	0.266
		oracle	0.120	0.198	0.338	0.304	0.530	0.288	0.264
		PWQEE.id	−0.540	0.333	−0.382	0.401	−0.345	0.276	0.337
		PWQEE.ex	0.520	0.150	−2.285	0.326	0.558	0.146	0.207
	$\alpha = 0.9$	PWQEE.ar	−0.361	0.115	−0.265	0.116	−0.112	0.090	0.107
		PQEE.ar	−0.282	0.105	−0.345	0.111	−0.079	0.081	0.099
		oracle	−0.351	0.116	−0.272	0.116	−0.096	0.091	0.108
		PWQEE.id	0.075	0.282	0.317	0.469	0.089	0.362	0.371
		PWQEE.ex	−0.199	0.216	0.329	0.345	0.091	0.280	0.280
$\tau = 0.75$	$\alpha = 0.5$	PWQEE.ar	−0.125	0.217	0.344	0.329	0.304	0.292	0.279
		PQEE.ar	−0.172	0.215	0.355	0.320	0.316	0.276	0.271
		oracle	−0.267	0.210	0.374	0.324	0.179	0.258	0.264
		PWQEE.id	−0.648	0.303	0.243	0.382	0.069	0.374	0.353
		PWQEE.ex	−0.637	0.305	0.097	0.399	0.084	0.357	0.353
	$\alpha = 0$	PWQEE.ar	−0.656	0.303	0.229	0.382	0.093	0.375	0.353
		PQEE.ar	−0.715	0.301	0.263	0.386	0.150	0.375	0.354
		oracle	−0.659	0.313	0.208	0.387	0.075	0.358	0.353

It can be seen that when the errors are multivariate normal (case 1, Tables 4 and Table 7) or multivariate t (case 2, Tables 5 and 8), and the correlation parameter α is relatively large, the parameter estimation of PWQEE method with AR(1) correlation structure is significantly better than that of PWQEE method with independence or exchangeable structure, which shows that the parameter estimates can be improved by correctly specifying the correlation structure. Similarly, when the data are not contaminated, the parameter estimate of PQEE.ar method is slightly better than that of the weighted method (PWQEE.ar). However, when there are outliers or high leverage points in the covariates (case 3, Tables 6 and 9), the parameter estimates can be significantly improved by the methods with weights, which

Table 5. Parameter estimation results of non-zero coefficients for case 2 when $p = 8$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
$\tau = 0.25$	$\alpha = 0.9$	PWQEE.id	−0.388	0.353	−0.518	0.503	0.415	0.418	0.424
		PWQEE.ex	0.526	0.197	−2.066	0.376	0.811	0.196	0.257
		PWQEE.ar	−0.227	0.138	0.002	0.182	−0.240	0.146	0.155
		PQEE.ar	−0.127	0.140	−0.071	0.172	−0.221	0.138	0.150
		oracle	−0.235	0.136	−0.034	0.185	−0.231	0.146	0.156
	$\alpha = 0.5$	PWQEE.id	−1.046	0.634	1.151	0.496	1.062	0.529	0.553
		PWQEE.ex	−0.801	0.527	0.787	0.405	0.880	0.426	0.453
		PWQEE.ar	−0.991	0.476	0.794	0.363	0.799	0.415	0.418
		PQEE.ar	−0.792	0.459	0.746	0.340	0.776	0.396	0.399
		oracle	−0.925	0.474	0.800	0.360	0.826	0.376	0.404
	$\alpha = 0$	PWQEE.id	0.543	0.414	−0.454	0.441	0.538	0.422	0.426
		PWQEE.ex	0.565	0.401	−0.396	0.431	0.744	0.431	0.421
		PWQEE.ar	0.572	0.427	−0.376	0.445	0.449	0.415	0.429
		PQEE.ar	0.654	0.429	−0.588	0.456	0.609	0.412	0.432
		oracle	0.623	0.402	−0.426	0.436	0.698	0.421	0.419
$\tau = 0.5$	$\alpha = 0.9$	PWQEE.id	0.002	0.305	0.395	0.524	−0.904	0.267	0.365
		PWQEE.ex	0.616	0.116	−1.434	0.283	0.389	0.105	0.168
		PWQEE.ar	−0.181	0.082	0.311	0.117	−0.355	0.086	0.095
		PQEE.ar	−0.259	0.083	0.351	0.116	−0.297	0.077	0.092
		oracle	−0.181	0.082	0.311	0.117	−0.355	0.086	0.095
	$\alpha = 0.5$	PWQEE.id	0.066	0.351	−0.388	0.392	−0.387	0.353	0.365
		PWQEE.ex	0.177	0.308	−0.431	0.315	−0.670	0.312	0.311
		PWQEE.ar	−0.133	0.300	−0.064	0.300	−0.664	0.304	0.301
		PQEE.ar	−0.157	0.289	−0.092	0.303	−0.573	0.300	0.297
		oracle	−0.126	0.301	−0.065	0.299	−0.685	0.303	0.301
	$\alpha = 0$	PWQEE.id	0.516	0.284	0.232	0.333	−0.047	0.295	0.304
		PWQEE.ex	0.459	0.286	0.233	0.329	0.042	0.293	0.303
		PWQEE.ar	0.469	0.295	0.209	0.330	0.002	0.296	0.307
		PQEE.ar	0.466	0.299	0.151	0.331	−0.041	0.297	0.309
		oracle	0.476	0.295	0.182	0.332	0.134	0.290	0.306
$\tau = 0.75$	$\alpha = 0.9$	PWQEE.id	−0.355	0.417	1.759	0.609	0.394	0.591	0.539
		PWQEE.ex	0.523	0.153	−0.981	0.381	1.386	0.233	0.256
		PWQEE.ar	−0.493	0.125	1.105	0.226	0.464	0.175	0.176
		PQEE.ar	−0.471	0.127	0.954	0.212	0.526	0.166	0.168
		oracle	−0.493	0.125	1.105	0.226	0.464	0.175	0.176
	$\alpha = 0.5$	PWQEE.id	−0.024	0.432	−0.327	0.394	1.435	0.460	0.429
		PWQEE.ex	0.138	0.337	−0.990	0.373	1.262	0.392	0.367
		PWQEE.ar	0.064	0.335	−0.561	0.325	1.377	0.377	0.346
		PQEE.ar	0.040	0.337	−0.603	0.327	1.385	0.372	0.345
		oracle	0.098	0.330	−0.558	0.323	1.016	0.354	0.336
	$\alpha = 0$	PWQEE.id	−0.027	0.438	−1.028	0.448	2.314	0.466	0.451
		PWQEE.ex	−0.122	0.435	−1.029	0.456	2.200	0.440	0.444
		PWQEE.ar	−0.129	0.432	−1.038	0.444	2.256	0.458	0.445
		PQEE.ar	−0.145	0.417	−1.050	0.440	2.315	0.458	0.438
		oracle	−0.142	0.432	−0.992	0.456	2.000	0.432	0.440

further suggests that adding weights can alleviate the influence of outliers in the covariates and improve the robustness of the estimation. When the dimension of the covariates increases from $p = 8$ to $p = 100$, biases and MSEs of all the methods are very close.

4. Real-data analysis

We now illustrate the proposed method via the yeast cell cycle gene expression data [36]. The whole-genome mRNA levels of 6178 yeast open reading frames were recorded

Table 6. Parameter estimation results of non-zero coefficients for case 3 when $p = 8$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
		PWQEE.id	−0.975	0.310	0.337	0.359	0.364	0.365	0.345
		PWQEE.ex	0.723	0.243	−1.855	0.344	0.999	0.317	0.301
	$\alpha = 0.9$	PWQEE.ar	−0.061	0.203	0.063	0.176	0.191	0.250	0.210
		PQEE.ar	−49.816	25.549	−25.204	7.429	−47.985	23.916	18.965
		oracle	−0.052	0.203	0.043	0.177	0.160	0.251	0.210
		PWQEE.id	−0.145	0.296	0.586	0.419	−0.949	0.454	0.389
		PWQEE.ex	0.393	0.280	−0.284	0.357	−0.326	0.346	0.328
$\tau = 0.25$	$\alpha = 0.5$	PWQEE.ar	0.139	0.253	0.626	0.347	−0.607	0.397	0.332
		PQEE.ar	−47.832	23.562	−23.918	6.586	−48.263	24.186	18.111
		oracle	0.155	0.263	0.614	0.350	−0.706	0.355	0.323
		PWQEE.id	−1.053	0.277	0.396	0.333	−1.288	0.396	0.335
		PWQEE.ex	−0.944	0.277	0.325	0.338	−1.291	0.352	0.322
	$\alpha = 0$	PWQEE.ar	−1.045	0.275	0.393	0.337	−1.274	0.397	0.336
		PQEE.ar	−48.106	23.712	−23.953	6.784	−48.827	24.403	18.300
		oracle	−0.944	0.276	0.353	0.337	−1.211	0.331	0.315
		PWQEE.id	−0.326	0.268	−0.323	0.287	−0.188	0.199	0.252
		PWQEE.ex	2.076	0.432	−2.929	0.403	1.657	0.312	0.382
	$\alpha = 0.9$	PWQEE.ar	0.722	0.273	−0.250	0.202	0.552	0.270	0.248
		PQEE.ar	−53.212	29.980	−38.963	28.192	−51.351	27.781	28.651
		oracle	0.722	0.273	−0.250	0.202	0.552	0.270	0.248
		PWQEE.id	−1.375	0.249	0.065	0.299	0.388	0.243	0.264
		PWQEE.ex	−0.935	0.231	−0.417	0.261	1.046	0.218	0.237
$\tau = 0.5$	$\alpha = 0.5$	PWQEE.ar	−1.027	0.220	−0.195	0.255	1.150	0.231	0.235
		PQEE.ar	−54.246	30.969	−40.706	31.046	−50.305	27.221	29.745
		oracle	−1.008	0.226	−0.202	0.256	1.175	0.235	0.239
		PWQEE.id	−0.826	0.269	0.542	0.400	−0.982	0.245	0.305
		PWQEE.ex	−0.763	0.266	0.493	0.377	−1.009	0.238	0.294
	$\alpha = 0$	PWQEE.ar	−0.847	0.268	0.537	0.396	−0.962	0.245	0.303
		PQEE.ar	−53.744	31.200	−39.390	30.151	−53.977	30.846	30.732
		oracle	−0.773	0.264	0.522	0.389	−1.006	0.232	0.295
		PWQEE.id	−0.964	0.323	0.471	0.464	−0.722	0.413	0.400
		PWQEE.ex	0.719	0.398	−2.849	0.546	1.156	0.367	0.437
	$\alpha = 0.9$	PWQEE.ar	−0.253	0.278	−0.337	0.283	0.196	0.305	0.289
		PQEE.ar	−27.534	7.944	−13.449	2.337	−26.298	7.332	5.871
		oracle	−0.246	0.283	−0.336	0.283	0.119	0.304	0.290
		PWQEE.id	−1.029	0.297	−0.100	0.438	0.716	0.375	0.370
		PWQEE.ex	−0.855	0.288	−0.733	0.438	1.079	0.283	0.336
$\tau = 0.75$	$\alpha = 0.5$	PWQEE.ar	−1.046	0.279	−0.521	0.389	0.902	0.306	0.325
		PQEE.ar	−27.826	8.089	−13.552	2.352	−25.764	6.979	5.807
		oracle	−1.143	0.287	−0.564	0.404	0.661	0.278	0.323
		PWQEE.id	−0.457	0.295	−0.392	0.372	0.354	0.316	0.328
		PWQEE.ex	−0.350	0.293	−0.504	0.359	0.195	0.274	0.309
	$\alpha = 0$	PWQEE.ar	−0.444	0.296	−0.395	0.374	0.369	0.317	0.329
		PQEE.ar	−26.127	7.142	−14.008	2.305	−26.049	7.080	5.509
		oracle	−0.357	0.289	−0.310	0.372	0.251	0.265	0.309

with a 7-min interval for 119 min, covering two cell cycles with a total of 18 time points.

The cell cycle is the series of events that take place in a cell that cause it to divide into two daughter cells. In this process, the genetic materials of the cell replicate and then flow into the two daughter cells. The cell cycle process is commonly divided into five stages: M/G1-G1-S-G2-M, where the M stage stands for 'mitosis ', during which the replicated chromosomes are separated into two new nuclei; G1 refers to 'Gap1' and cells increase in size in this stage; S stands for 'synthesis' during which DNA replication occurs; G2 refers

Table 7. Parameter estimation results of non-zero coefficients for case 1 when $p = 100$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
		PWQEE.id	0.174	0.242	−0.961	0.408	0.460	0.321	0.324
		PWQEE.ex	0.847	0.159	−3.120	0.410	1.270	0.176	0.248
	$\alpha = 0.9$	PWQEE.ar	−0.058	0.132	−0.478	0.172	0.270	0.127	0.144
		PQEE.ar	−0.071	0.126	−0.436	0.149	0.117	0.122	0.132
		oracle	−0.058	0.132	−0.478	0.172	0.270	0.127	0.144
		PWQEE.id	0.151	0.288	0.605	0.518	−1.001	0.338	0.381
		PWQEE.ex	0.160	0.244	0.109	0.431	−0.755	0.280	0.318
$\tau = 0.25$	$\alpha = 0.5$	PWQEE.ar	−0.004	0.248	0.343	0.400	−1.007	0.259	0.302
		PQEE.ar	0.094	0.239	0.341	0.394	−0.990	0.256	0.296
		oracle	−0.152	0.233	0.492	0.381	−0.916	0.250	0.288
		PWQEE.id	0.494	0.327	0.438	0.338	−0.491	0.275	0.313
		PWQEE.ex	0.536	0.346	0.256	0.342	−0.509	0.274	0.321
	$\alpha = 0$	PWQEE.ar	0.501	0.311	0.418	0.326	−0.422	0.277	0.304
		PQEE.ar	0.416	0.298	0.209	0.329	−0.630	0.273	0.300
		oracle	0.483	0.349	0.420	0.335	−0.558	0.274	0.320
		PWQEE.id	−0.328	0.248	0.204	0.336	−0.606	0.167	0.250
		PWQEE.ex	0.502	0.119	−1.781	0.218	0.316	0.092	0.143
	$\alpha = 0.9$	PWQEE.ar	−0.060	0.083	−0.287	0.099	−0.232	0.079	0.087
		PQEE.ar	−0.169	0.079	−0.252	0.105	−0.217	0.071	0.085
		oracle	−0.064	0.083	−0.283	0.099	−0.247	0.079	0.087
		PWQEE.id	−0.013	0.259	0.823	0.303	−0.555	0.232	0.265
		PWQEE.ex	0.137	0.185	0.494	0.252	−0.435	0.187	0.208
$\tau = 0.5$	$\alpha = 0.5$	PWQEE.ar	0.077	0.183	0.597	0.260	−0.484	0.182	0.208
		PQEE.ar	0.157	0.175	0.628	0.249	−0.524	0.169	0.197
		oracle	0.111	0.183	0.557	0.261	−0.502	0.182	0.209
		PWQEE.id	−0.080	0.266	0.531	0.258	0.581	0.187	0.237
		PWQEE.ex	−0.016	0.262	0.388	0.242	0.602	0.186	0.230
	$\alpha = 0$	PWQEE.ar	−0.023	0.269	0.502	0.255	0.605	0.185	0.236
		PQEE.ar	0.016	0.262	0.370	0.239	0.579	0.176	0.226
		oracle	−0.041	0.264	0.440	0.252	0.570	0.188	0.235
		PWQEE.id	0.507	0.229	−0.794	0.349	0.416	0.264	0.281
		PWQEE.ex	1.552	0.161	−3.179	0.462	1.465	0.191	0.271
	$\alpha = 0.9$	PWQEE.ar	0.131	0.100	−0.172	0.134	0.191	0.094	0.109
		PQEE.ar	0.124	0.111	−0.191	0.130	0.223	0.095	0.112
		oracle	0.131	0.100	−0.172	0.134	0.191	0.094	0.109
		PWQEE.id	−0.774	0.266	0.822	0.348	−0.014	0.284	0.300
		PWQEE.ex	−0.734	0.260	0.453	0.338	0.207	0.243	0.281
$\tau = 0.75$	$\alpha = 0.5$	PWQEE.ar	−0.904	0.234	1.019	0.333	−0.030	0.226	0.264
		PQEE.ar	−1.040	0.225	1.027	0.293	0.027	0.212	0.243
		oracle	−0.973	0.243	1.185	0.326	−0.128	0.222	0.264
		PWQEE.id	0.118	0.303	0.192	0.458	−0.234	0.403	0.388
		PWQEE.ex	0.252	0.309	−0.483	0.464	0.167	0.373	0.382
	$\alpha = 0$	PWQEE.ar	0.104	0.299	0.104	0.457	−0.318	0.392	0.383
		PQEE.ar	−0.145	0.308	0.095	0.440	0.009	0.385	0.378
		oracle	0.055	0.300	0.001	0.453	−0.034	0.365	0.373

to ‘Gap2’, during this stage, the cell will continue to grow and the G2 checkpoint control mechanism will ensure that everything is ready to enter the M phase and divide. Spellman et al. [36] identified a total of about 800 genes showing periodic patterns during the yeast cell cycle, but the regulation of most of these genes remains unknown. In this study, we apply the proposed PWQEE method to identify the transcription factors (TFs) that highly influence the gene expression at each stage of the yeast cell cycle under different quantile levels, which can facilitate the understanding of the regulation mechanism of cell cycles.

Table 8. Parameter estimation results of non-zero coefficients for case 2 when $p = 100$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
$\tau = 0.25$	$\alpha = 0.9$	PWQEE.id	−0.513	0.469	0.928	0.535	−0.417	0.369	0.458
		PWQEE.ex	1.254	0.260	−3.164	0.517	0.983	0.219	0.333
		PWQEE.ar	−0.165	0.183	0.415	0.190	−0.303	0.146	0.173
		PQEE.ar	−0.014	0.165	0.364	0.188	−0.428	0.132	0.173
		oracle	−0.183	0.183	0.429	0.190	−0.304	0.146	0.156
	$\alpha = 0.5$	PWQEE.id	0.718	0.397	−0.419	0.422	0.993	0.476	0.431
		PWQEE.ex	0.926	0.349	−0.640	0.437	0.660	0.404	0.396
		PWQEE.ar	0.589	0.327	0.009	0.432	0.616	0.414	0.391
		PQEE.ar	0.431	0.317	−0.014	0.428	0.595	0.410	0.385
		oracle	0.483	0.340	0.001	0.415	0.496	0.395	0.383
	$\alpha = 0$	PWQEE.id	0.659	0.510	0.189	0.696	−0.203	0.455	0.553
		PWQEE.ex	0.902	0.523	−0.478	0.699	−0.181	0.437	0.553
		PWQEE.ar	0.633	0.508	0.201	0.702	−0.303	0.467	0.559
		PQEE.ar	0.480	0.503	0.404	0.664	−0.451	0.457	0.541
		oracle	0.705	0.507	0.017	0.678	−0.459	0.442	0.542
$\tau = 0.5$	$\alpha = 0.9$	PWQEE.id	−0.510	0.260	0.386	0.312	0.170	0.259	0.277
		PWQEE.ex	0.689	0.130	−1.911	0.285	1.062	0.129	0.181
		PWQEE.ar	−0.269	0.084	0.376	0.119	−0.009	0.089	0.098
		PQEE.ar	−0.207	0.079	0.214	0.134	0.094	0.088	0.100
		oracle	−0.269	0.084	0.376	0.119	−0.009	0.088	0.098
	$\alpha = 0.5$	PWQEE.id	0.173	0.294	−0.189	0.345	0.003	0.281	0.307
		PWQEE.ex	0.113	0.249	−0.359	0.294	0.031	0.220	0.255
		PWQEE.ar	0.185	0.226	−0.341	0.278	0.135	0.242	0.248
		PQEE.ar	0.226	0.216	−0.379	0.254	0.202	0.249	0.240
		oracle	0.155	0.224	−0.344	0.278	0.143	0.242	0.248
	$\alpha = 0$	PWQEE.id	−0.138	0.321	−0.178	0.354	−0.848	0.307	0.327
		PWQEE.ex	−0.140	0.324	−0.351	0.351	−0.840	0.304	0.326
		PWQEE.ar	−0.175	0.324	−0.192	0.361	−0.883	0.311	0.332
		PQEE.ar	−0.293	0.323	−0.182	0.350	−0.775	0.286	0.319
		oracle	−0.226	0.318	−0.145	0.352	−0.929	0.308	0.326
$\tau = 0.75$	$\alpha = 0.9$	PWQEE.id	−1.186	0.411	1.539	0.580	0.792	0.512	0.501
		PWQEE.ex	0.677	0.191	−2.280	0.377	1.661	0.250	0.272
		PWQEE.ar	−0.269	0.133	−0.008	0.193	0.863	0.189	0.172
		PQEE.ar	−0.219	0.128	0.024	0.181	0.686	0.181	0.163
		oracle	−0.286	0.132	−0.094	0.193	0.854	0.189	0.172
	$\alpha = 0.5$	PWQEE.id	0.184	0.539	−0.186	0.399	0.477	0.323	0.420
		PWQEE.ex	0.099	0.442	−0.644	0.367	0.657	0.287	0.365
		PWQEE.ar	0.005	0.419	−0.049	0.353	0.099	0.296	0.356
		PQEE.ar	−0.085	0.391	0.063	0.335	0.213	0.289	0.338
		oracle	−0.188	0.429	0.006	0.359	0.250	0.270	0.353
	$\alpha = 0$	PWQEE.id	0.305	0.360	−0.569	0.473	−0.438	0.384	0.406
		PWQEE.ex	0.507	0.352	−0.819	0.437	−0.375	0.389	0.393
		PWQEE.ar	0.297	0.364	−0.408	0.478	−0.413	0.391	0.411
		PQEE.ar	0.211	0.349	−0.453	0.489	−0.745	0.401	0.413
		oracle	0.358	0.364	−0.439	0.466	−0.443	0.406	0.412

Assume the data are missing completely at random, and we analyse 297 cell cycle regulatory genes (a total of 240 genes). The changes of their expression levels with time in cell cycles are shown in Figure 2. The response variable y_{ij} is the log-transformed gene expression level of gene i measured at t_{ij}; the covariate x_{ik} ($k = 1, \ldots, 96$) is the matching score of the binding probability of the kth TF on the promoter region of the ith gene [37]. Similarly as Ref. [20], we apply PWQEE to the five stages of the cell cycle process based on the

Table 9. Parameter estimation results of non-zero coefficients for case 3 when $p = 100$.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		
			Bias	MSE	Bias	MSE	Bias	MSE	AM
$\tau = 0.25$	$\alpha = 0.9$	PWQEE.id	−0.046	0.255	−0.817	0.386	0.179	0.326	0.322
		PWQEE.ex	2.582	0.458	−4.912	0.786	1.743	0.464	0.569
		PWQEE.ar	0.587	0.354	0.549	0.260	0.106	0.413	0.342
		PQEE.ar	−22.634	6.016	−3.236	1.130	−16.770	3.834	3.660
		oracle	0.586	0.354	0.549	0.260	0.106	0.414	0.343
	$\alpha = 0.5$	PWQEE.id	−0.662	0.326	−0.135	0.425	−0.636	0.318	0.356
		PWQEE.ex	−0.020	0.352	−1.358	0.472	0.183	0.369	0.398
		PWQEE.ar	−0.624	0.316	0.049	0.432	0.313	0.360	0.369
		PQEE.ar	−22.876	6.076	−6.346	1.598	−14.039	3.006	3.560
		oracle	−0.577	0.320	0.060	0.436	−0.390	0.363	0.373
	$\alpha = 0$	PWQEE.id	−0.383	0.376	−0.153	0.366	0.235	0.484	0.409
		PWQEE.ex	−0.422	0.362	−0.286	0.348	0.445	0.434	0.381
		PWQEE.ar	−0.421	0.385	−0.077	0.364	0.240	0.489	0.413
		PQEE.ar	−21.494	5.459	−8.046	1.685	−13.544	2.798	3.314
		oracle	−0.574	0.373	0.009	0.387	0.303	0.443	0.402
$\tau = 0.5$	$\alpha = 0.9$	PWQEE.id	0.441	0.234	0.367	0.281	−0.707	0.266	0.260
		PWQEE.ex	1.901	0.292	−3.887	0.568	1.226	0.293	0.384
		PWQEE.ar	0.414	0.218	0.125	0.247	−0.336	0.298	0.254
		PQEE.ar	−52.570	28.509	−26.527	7.799	−54.442	30.469	22.259
		oracle	0.414	0.218	0.125	0.247	−0.336	0.298	0.255
	$\alpha = 0.5$	PWQEE.id	−0.520	0.247	−0.010	0.324	0.377	0.299	0.290
		PWQEE.ex	−0.423	0.241	−0.531	0.330	0.390	0.246	0.272
		PWQEE.ar	−0.699	0.250	−0.161	0.276	0.191	0.228	0.251
		PQEE.ar	−53.673	29.569	−26.507	7.884	−53.998	29.976	22.476
		oracle	−0.701	0.249	−0.144	0.276	0.207	0.227	0.251
	$\alpha = 0$	PWQEE.id	−0.194	0.296	−0.042	0.377	−0.381	0.311	0.328
		PWQEE.ex	−0.127	0.291	−0.252	0.371	−0.343	0.316	0.326
		PWQEE.ar	−0.233	0.298	−0.005	0.378	−0.376	0.309	0.328
		PQEE.ar	−54.279	30.047	−25.714	7.357	−54.855	30.821	22.742
		oracle	−0.168	0.296	−0.034	0.376	−0.431	0.309	0.327
$\tau = 0.75$	$\alpha = 0.9$	PWQEE.id	−0.978	0.291	1.064	0.390	−0.074	0.336	0.339
		PWQEE.ex	1.107	0.338	−3.773	0.629	1.571	0.420	0.463
		PWQEE.ar	−0.534	0.274	0.604	0.195	−0.079	0.311	0.260
		PQEE.ar	−27.498	7.909	−12.829	2.036	−26.404	7.282	5.742
		oracle	−0.534	0.274	0.604	0.195	−0.079	0.311	0.260
	$\alpha = 0.5$	PWQEE.id	0.148	0.333	−0.444	0.371	−0.626	0.370	0.358
		PWQEE.ex	0.077	0.316	−1.638	0.357	−0.362	0.340	0.338
		PWQEE.ar	−0.508	0.300	−0.625	0.300	−0.947	0.360	0.320
		PQEE.ar	−26.434	7.258	−13.717	2.206	−25.886	7.004	5.489
		oracle	−0.557	0.294	−0.593	0.288	−0.903	0.358	0.313
	$\alpha = 0$	PWQEE.id	−0.076	0.327	0.006	0.336	0.235	0.298	0.320
		PWQEE.ex	−0.002	0.322	−0.167	0.340	0.154	0.261	0.308
		PWQEE.ar	−0.036	0.323	−0.009	0.334	0.194	0.305	0.320
		PQEE.ar	−26.181	7.159	−13.047	2.075	−26.111	7.093	5.442
		oracle	−0.131	0.337	0.094	0.354	0.014	0.257	0.316

following model:

$$y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \sum_{k=1}^{96} \beta_k x_{ik} + \epsilon_{ij}, \tag{16}$$

where x_{ik} is the matching score after standardization; t_{ij} denotes time. We use 85% of the points in the data as the training set and the remaining 15% as the testing set to test the error of the model. Table 10 summarizes the numbers of TFs identified under three different

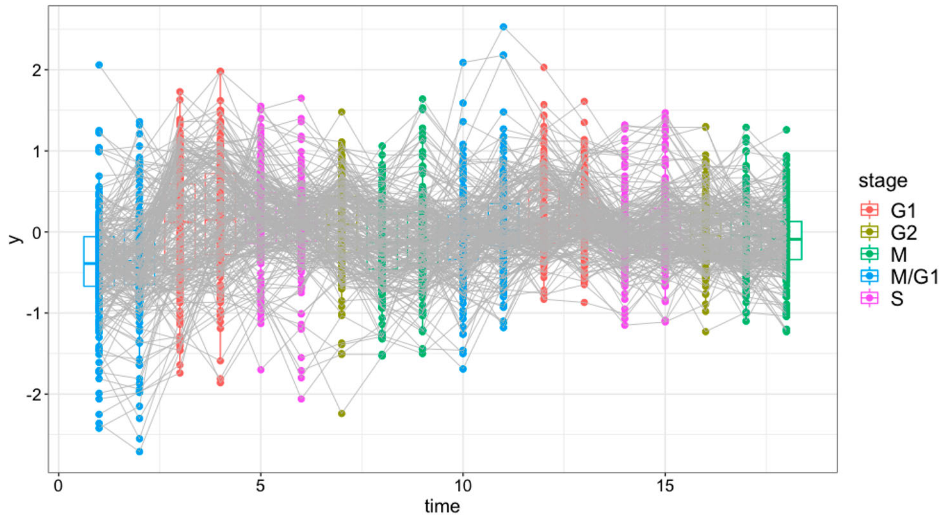

Figure 2. The changes of expression levels of 240 genes with time in two cell cycles.

Table 10. The number of TFs selected and test error for each stage in the yeast cell cycle process.

		Num			Test error		
		$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
M/G1	PWQEE.id	14	14	14	0.783	0.624	0.362
	PWQEE.ex	12	19	37	0.852	0.620	0.332
	PWQEE.ar	27	21	27	0.802	0.628	0.342
G1	PWQEE.id	24	20	24	0.590	0.190	0.180
	PWQEE.ex	12	12	21	0.463	0.223	0.181
	PWQEE.ar	20	23	23	0.577	0.178	0.171
S	PWQEE.id	47	14	33	0.202	0.399	0.545
	PWQEE.ex	33	12	25	0.209	0.375	0.535
	PWQEE.ar	24	20	30	0.183	0.373	0.562
G2	PWQEE.id	23	43	41	0.145	0.276	0.380
	PWQEE.ex	13	10	55	0.105	0.255	0.385
	PWQEE.ar	20	12	36	0.140	0.261	0.357
M	PWQEE.id	37	16	7	0.173	0.146	0.123
	PWQEE.ex	34	14	6	0.187	0.148	0.123
	PWQEE.ar	39	19	11	0.123	0.164	0.136

Note: Num: the numbers of TFs identified.

working correlation structures (independence, exchangeable and AR(1)) and the test error based on the check function.

We also give the coefficients of the 96 TFs at five stages under the AR(1) correlation structure, as shown in Figure 3. There is a large proportion of overlap between the TFs selected by our method and the results identified in Table 2 of Ref. [16]. Some of the selected TFs have been confirmed by biological experiments. For example, FKH1, FKH2 and MCM1 are the three TFs that have been proved to be important for the G2 phase according to the above biological experiments. Our method successfully selects them in the G2 phase. In addition, similar to the results of Wang et al. [20], only a small part of TFs selected at different stages of the cell cycle overlaps. We also find that some selected

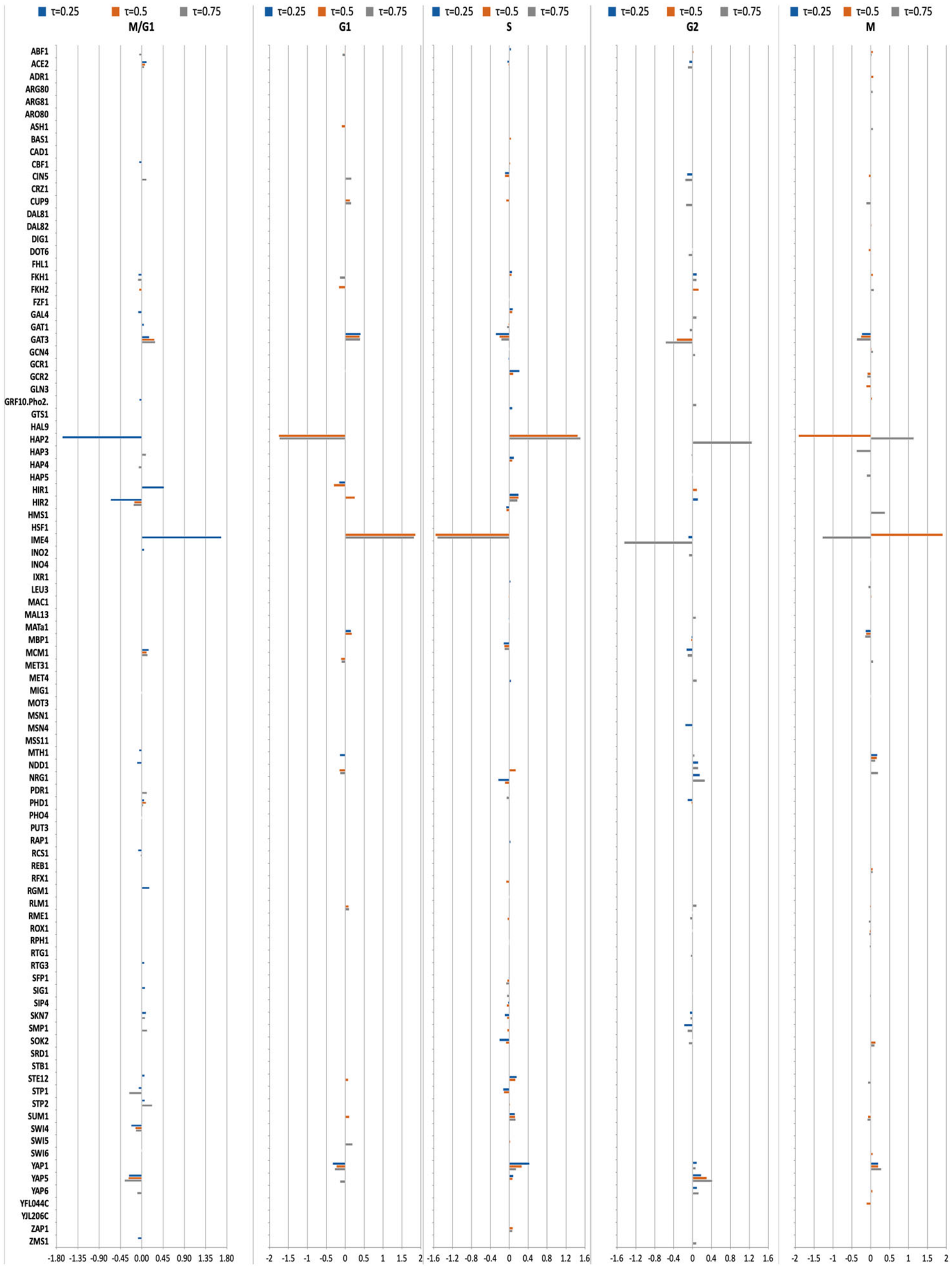

Figure 3. The coefficients derived from the PWQEEs with AR(1) correlation structure at five stages in yeast cell cycle.

TFs have opposite effects on gene expression during different stages. For example, HAP2 inhibits gene expression in the G1 phase, but promotes gene expression in the S phase; Gat3 promotes gene expression in the G1 phase but turns to inhibition in the S phase. At the same time, many selected TFs have different effects on the gene expression level under different quantiles. For instance, HAP2 significantly affects the gene expression under the

75th quantile during the G2 stage. However, under the 25th and 50th quantiles, its effect is negligible. All of these observations indicate that different TFs play important roles in the regulation of genes at different stages of the cell cycle, which has also been observed by biologists.

5. Conclusions

We proposed a variable selection method PWQEE, which can give sparse, robust and efficient estimators for quantile models with longitudinal data. Instead of assuming independence and minimizing the penalized quantile objective function, our method incorporated the working correlation structures to make use of the correlation information in longitudinal data. Furthermore, we added weights to the covariates to mitigate the impacts of high leverage points and outliers, which makes our method have double-robustness to the non-normal distributed errors and the contaminated covariates. The simulation showed that the PWQEE outperforms in variable selection and parameter estimation when dealing with the correlated data. The proposed method can be extended to the nonparametric QR models, which we will consider in future work.

Acknowledgments

The authors thank the Associate Editor and referees for their constructive comments. The authors acknowledge the technical support by the HPC platform in Xi'an Jiaotong University.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author L. Y. Fu upon reasonable request.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China (No. 11871390) and Shaanxi Fundamental Science Research Project for Mathematics and Physics (No. 22JSY002).

References

[1] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B. 1996;58(1):267–288.
[2] Frank LE, Friedman JH. A statistical view of some chemometrics regression tools. Technometrics. 1993;35(2):109–135.
[3] Fan JQ, Li RZ. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348–1360.
[4] Zou H. The adaptive lasso and its oracle properties. J Am Stat Assoc. 2006;101(476):1418–1429.
[5] Hunter DR, Li RZ. Variable selection using MM algorithms. Ann Stat. 2005;33(4):1617–1642.
[6] Koenker R, Bassett Jr G. Regression quantiles. Econ J Econ Soc. 1978;46(1):33–50.
[7] Wu YC, Liu Y. Variable selection in quantile regression. Stat Sin. 2009;19(2):801–817.
[8] Zou H, Li RZ. One-step sparse estimates in nonconcave penalized likelihood models. Ann Stat. 2008;36(4):1509–1533.

[9] Peng B, Wang L. An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. J Comput Graph Stat. 2015;24(3):676–694.
[10] Fan JQ, Fan YY, Barut E. Adaptive robust variable selection. Ann Stat. 2014;42(1):324–351.
[11] Kato K. Estimation in functional linear quantile regression. Ann Stat. 2012;40(6):3108–3136.
[12] Ma H, Li T, Zhu H, et al. Quantile regression for functional partially linear model in ultra-high dimensions. Comput Stat Data Anal. 2019;129:135–147.
[13] Zhang YY, Lian H, Li GD, et al. Function additive quantile regression. Stat Sin. 2021;31:1–21.
[14] Diggle PJ, Heagerty P, Liang K-Y, et al. Analysis of longitudinal data. Oxford: Oxford University Press; 2002.
[15] Fan JQ, Li RZ. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J Am Stat Assoc. 2004;99(467):710–723.
[16] Wang LF, Li HZ, Huang JH. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. J Am Stat Assoc. 2008;103(484):1556–1569.
[17] Ni X, Zhang DW, Zhang HH. Variable selection for semiparametric mixed models in longitudinal studies. Biometrics. 2010;66(1):79–88.
[18] Xu P, Wu P, Wang Y, et al. A GEE based shrinkage estimation for the generalized linear model in longitudinal data analysis. Hong Kong: Department of Mathematics, Hong Kong Baptist University; 2010.
[19] Johnson BA, Lin DY, Zeng DL. Penalized estimating functions and variable selection in semiparametric regression models. J Am Stat Assoc. 2008;103(482):672–680.
[20] Wang L, Zhou JH, Qu A. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. Biometrics. 2012;68(2):353–360.
[21] Koenker R. Quantile regression for longitudinal data. J Multivar Anal. 2004;91(1):74–89.
[22] Wang HJ, Fygenson M. Inference for censored quantile regression models in longitudinal studies. Ann Stat. 2009;37(2):756–781.
[23] Leng C, Zhang W. Smoothing combined estimating equations in quantile regression for longitudinal data. Stat Comput. 2014;24(1):123–136.
[24] Zhang Y, Wang H, Zhu Z. Quantile-regression-based clustering for panel data. J Econ. 2019;213(1):54–67.
[25] Gao XL, Liu Q. Sparsity identification in ultra-high dimensional quantile regression models with longitudinal data. Commun Stat Theory Methods. 2020;49(19):4712–4736.
[26] Fu LY, Wang Y-G. Quantile regression for longitudinal data with a working correlation model. Comput Stat Data Anal. 2012;56(8):2526–2538.
[27] Wang N. Marginal nonparametric kernel regression accounting for within-subject correlation. Biometrika. 2003;90(1):43–52.
[28] Zhu ZY, Fung WK, He XM. On the asymptotic of marginal regression splines with longitudinal data. Biometrika. 2008;95(4):907–917.
[29] Chen L, Wei L-J, Parzen MI. Quantile regression for correlated observations. In: Proceedings of the Second Seattle Symposium in Biostatistics; 2004; Seattle. p. 51–69.
[30] Terpstra JT, McKean JW. Rank-based analysis of linear models using R. J Stat Softw. 2005;14(7):1–26.
[31] Rousseeuw PJ, Driessen KV. A fast algorithm for the minimum covariance determinant estimator. Technometrics. 1999;41(3):212–223.
[32] Fu LY, Wang Y-G. Efficient parameter estimation via Gaussian copulas for quantile regression with longitudinal data. J Multivar Anal. 2016;143:492–502.
[33] Wang Y-G, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika. 2003;90(1):29–41.
[34] Heller G. Smoothed rank regression with censored data. J Am Stat Assoc. 2007;102(478): s552–559.
[35] Wang HS, Li B, Leng C. Shrinkage tuning parameter selection with a diverging number of parameters. J R Stat Soc Ser B. 2009;71(3):671–683.

[36] Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle–regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell. 1998;9(12):3273–3297.
[37] Wang LF, Chen G, Li HZ. Group SCAD regression analysis for microarray time course gene expression data. Bioinformatics. 2007;23(12):1486–1494.
[38] Pollard D. Empirical processes: theory and applications. Hayward, CA: Institute of Mathematical Statistics; 1990.
[39] Jung S-H. Quasi-likelihood for median regression models. J Am Stat Assoc. 1996;91(433): 251–257.

Appendix

Proof of Lemma 2.1: Denote $\mathbf{G}_i^{\mathrm{T}} = \mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{\Omega}_i$, then $\mathbf{W}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}\mathbf{S}_i$. Let $\overline{\mathbf{W}}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}\mathbf{P}_i$, where $\mathbf{P}_i = (\tau - \mathbb{P}(y_{i1} - \mathbf{x}_{i1}^{\mathrm{T}}\boldsymbol{\beta} \leq 0), \ldots, \tau - \mathbb{P}(y_{in_i} - \mathbf{x}_{in_i}^{\mathrm{T}}\boldsymbol{\beta} \leq 0))^{\mathrm{T}}$, and we have

$$
\begin{aligned}
N^{-1}\left\{\mathbf{W}(\boldsymbol{\beta}) - \overline{\mathbf{W}}(\boldsymbol{\beta})\right\} &= N^{-1}\sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}\left(\mathbf{S}_i - \mathbf{P}_i\right) \\
&= N^{-1}\sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}} \begin{pmatrix} \mathbb{P}\left(y_{i1} - \mathbf{x}_{i1}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) - I\left(y_{i1} - \mathbf{x}_{i1}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) \\ \vdots \\ \mathbb{P}\left(y_{in_i} - \mathbf{x}_{in_i}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) - I\left(y_{in_i} - \mathbf{x}_{in_i}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) \end{pmatrix} \\
&= N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathbf{g}_{ij}\left[\mathbb{P}\left(y_{ij} - \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) - I\left(y_{ij} - \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right)\right],
\end{aligned}
$$

where $\mathbf{g}_{ij}$ is the jth column of $\mathbf{G}_i^{\mathrm{T}}$. Using the uniform strong law of large numbers together with condition A3 [38], we can obtain

$$
\sup_{\boldsymbol{\beta}\in\mathbb{B}}\left|N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathbf{g}_{ij}\left[\mathbb{P}\left(y_{ij} - \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right) - I\left(y_{ij} - \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} \leq 0\right)\right]\right| \to o\left(N^{-1/2+\epsilon}\right) \quad \text{a.s.}
$$

Therefore,

$$
\sup_{\boldsymbol{\beta}\in\mathbb{B}} \|N^{-1}\left\{\mathbf{W}(\boldsymbol{\beta}) - \overline{\mathbf{W}}(\boldsymbol{\beta})\right\}\| = o(N^{-1/2+\epsilon}) \quad \text{a.s.} \tag{A1}
$$

Because $\mathbf{S}_i$ are random variables with mean zero, and the variance of $\mathbf{S}_i$ $\mathrm{Var}(N^{-1/2}\mathbf{W}(\boldsymbol{\beta}_0)) = (1/N)\sum_{i=1}^{N} \mathbf{X}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{\Omega}_i\mathbf{V}_i\mathbf{\Omega}_i\mathbf{V}_i^{-1}\mathbf{X}_i$ tends to $\mathbf{\Xi}$ with probability 1, the multivariate central limit theorem implies that $N^{-1/2}\mathbf{W}(\boldsymbol{\beta}_0) \to N(\mathbf{0}, \mathbf{\Xi})$.

For any $\boldsymbol{\beta}$ satisfying $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq KN^{-1/2}$ (note that in Jung's work [39], the proof is derived for any $\boldsymbol{\beta}$ satisfying $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq KN^{-1/3}$, here Jung's derivation still works because $\{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq KN^{-1/2}\} \subset \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq KN^{-1/3}\}$),

$$
\begin{aligned}
\mathbf{W}(\boldsymbol{\beta}) - \mathbf{W}\left(\boldsymbol{\beta}_0\right) &= \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}(\boldsymbol{\beta})\mathbf{S}_i(\boldsymbol{\beta}) - \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_0\right)\mathbf{S}_i\left(\boldsymbol{\beta}_0\right) \\
&= \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}(\boldsymbol{\beta})\left\{\mathbf{S}_i(\boldsymbol{\beta}) - \mathbf{S}_i\left(\boldsymbol{\beta}_0\right)\right\} + \sum_{i=1}^{N}\left\{\mathbf{G}_i(\boldsymbol{\beta}) - \mathbf{G}_i\left(\boldsymbol{\beta}_0\right)\right\}^{\mathrm{T}}\mathbf{S}_i\left(\boldsymbol{\beta}_0\right).
\end{aligned} \tag{A2}
$$

According to the lemma in Ref. [39], the first term of (A2)

$$
\sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}(\boldsymbol{\beta})\left\{\mathbf{S}_i(\boldsymbol{\beta}) - \mathbf{S}_i\left(\boldsymbol{\beta}_0\right)\right\} = \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}(\boldsymbol{\beta})\mathbf{P}_i(\boldsymbol{\beta}) + \sum_{i=1}^{N} \mathbf{G}_i^{\mathrm{T}}(\boldsymbol{\beta})\left\{\mathbf{S}_i(\boldsymbol{\beta}) - \mathbf{S}_i\left(\boldsymbol{\beta}_0\right) - \mathbf{P}_i(\boldsymbol{\beta})\right\}
$$

$$= \overline{\boldsymbol{W}}(\boldsymbol{\beta}) + o_p\left(N^{1/2}\right).$$

Because $\mathbb{P}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta_0} \le 0) = \tau$, the second term of (A2)

$$\sum_{i=1}^{N} \{\boldsymbol{G_i}(\boldsymbol{\beta}) - \boldsymbol{G_i}(\boldsymbol{\beta_0})\}^{\mathrm{T}} \boldsymbol{S_i}(\boldsymbol{\beta_0}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \left(\boldsymbol{g_{ij}}(\boldsymbol{\beta}) - \boldsymbol{g_{ij}}(\boldsymbol{\beta_0})\right)$$
$$\times \left[\mathbb{P}\left(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta_0} \le 0\right) - I\left(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta_0} \le 0\right)\right].$$

According to Pollard [38], $\sum_{i=1}^{N}\{\boldsymbol{G_i}(\boldsymbol{\beta}) - \boldsymbol{G_i}(\boldsymbol{\beta_0})\}^{\mathrm{T}}\boldsymbol{S_i}(\boldsymbol{\beta_0}) = o_p(N^{1/2+\epsilon})$. Hence,

$$\boldsymbol{W}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta_0}) = \overline{\boldsymbol{W}}(\boldsymbol{\beta}) + o_p(N^{1/2}).$$

Note that $\overline{\boldsymbol{W}}(\boldsymbol{\beta_0}) = \boldsymbol{0}$. By Taylor's expansion of $\overline{\boldsymbol{W}}(\boldsymbol{\beta})$, we have

$$N^{-1/2}\{\boldsymbol{W}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta_0})\} = \frac{1}{N}\frac{\partial \overline{\boldsymbol{W}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}} N^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta_0}) + o_p(1).$$

By calculation we can get $N^{-1}(\partial\overline{\boldsymbol{W}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{\Omega_i}\boldsymbol{\Lambda_i}\boldsymbol{X_i}$, which is positive definite under condition A3. By taking the limitation, we obtain that

$$\sup_{||\boldsymbol{\beta}-\boldsymbol{\beta_0}||\le KN^{-1/2}} ||N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta}) - N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta_0}) - N^{1/2}\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta_0})|| = o_p(1).$$

This finishes the proof of Lemma 2.1. Furthermore, if we denote the estimator derived from solving $\boldsymbol{W}(\boldsymbol{\beta}) = \boldsymbol{0}$ by $\widehat{\boldsymbol{\beta}}_W$, the consistency and normality of $\widehat{\boldsymbol{\beta}}_W$ can be easily established. Since $\boldsymbol{\beta_0}$ is the unique solution to $\overline{\boldsymbol{W}}(\boldsymbol{\beta})$. This, together with (A1) and the definition of $\widehat{\boldsymbol{\beta}}_W$ implies that $\widehat{\boldsymbol{\beta}}_W \to \boldsymbol{\beta_0}$ as $N \to \infty$. Because $\boldsymbol{W}(\widehat{\boldsymbol{\beta}}_W) = 0$ and $\widehat{\boldsymbol{\beta}}_W$ is in the $N^{-1/2}$ neighbourhood of $\boldsymbol{\beta_0}$, we have

$$N^{1/2}(\widehat{\boldsymbol{\beta}}_W - \boldsymbol{\beta_0}) = -\left[N^{-1}\frac{\partial\overline{\boldsymbol{W}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}}\right]^{-1} N^{-1/2}\boldsymbol{W}(\boldsymbol{\beta_0}) + o_p(1).$$

Therefore, $N^{1/2}(\widehat{\boldsymbol{\beta}}_W - \boldsymbol{\beta_0}) \to N(\boldsymbol{0}, \boldsymbol{\Xi}')$, where

$$\boldsymbol{\Xi}' = \lim_{N\to\infty}\left[N^{-1}\frac{\partial\overline{\boldsymbol{W}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}}\right]^{-1}\boldsymbol{\Xi}\left\{\left[N^{-1}\frac{\partial\overline{\boldsymbol{W}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta_0}}\right]^{-1}\right\}^{\mathrm{T}}.$$

■

Proof of Lemma 2.2: By direct calculation, we have $\tilde{S}_{ij} - S_{ij} = \mathrm{sgn}(-\epsilon_{ij})\Phi(-|\epsilon_{ij}/h|)$. Because the bandwidth $h > 0$, we can write $\tilde{S}_{ij} - S_{ij} = \mathrm{sgn}(-d_{ij})\Phi(-|d_{ij}|)$, where $d_{ij} = \epsilon_{ij}/h$. We have

$$N^{-1/2}\left\{\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta})\right\} = N^{-1/2}\sum_{i=1}^{N}\boldsymbol{G}_i^{\mathrm{T}}\begin{pmatrix}\mathrm{sgn}(-d_{i1})\,\Phi(-|d_{i1}|)\\ \vdots \\ \mathrm{sgn}(-d_{in_i})\,\Phi(-|d_{in_i}|)\end{pmatrix}$$
$$= N^{-1/2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\boldsymbol{g_{ij}}\,\mathrm{sgn}(-d_{ij})\,\Phi(-|d_{ij}|),$$

where $\boldsymbol{G}_i^{\mathrm{T}} = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{\Omega_i}$, and $\boldsymbol{g_{ij}}$ is the jth column of $\boldsymbol{G}_i^{\mathrm{T}}$. Then focusing on the expectation of this difference,

$$\mathbb{E}\left(\tilde{S}_{ij} - S_{ij}\right) = \int_{-\infty}^{+\infty}\mathrm{sgn}(-d_{ij})\,\Phi(-|d_{ij}|)\,f_{ij}(\epsilon)\,\mathrm{d}\epsilon$$

$$= \int_{-\infty}^{+\infty} \Phi\left(-|\epsilon|/h\right)\{2I(\epsilon \le 0) - 1\} f_{ij}(\epsilon)\, \mathrm{d}\epsilon$$

$$= h \int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\} \left[f_{ij}(0) + f'_{ij}(\zeta(t)) ht \right] \mathrm{d}t,$$

where $\zeta(t)$ is between 0 and ht. Because $\int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\}\, \mathrm{d}t = 0$, we have $h \int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\} f_{ij}(0)\, \mathrm{d}t = 0$. Since $\int_{-\infty}^{+\infty} |t|\Phi(-|t|)\, \mathrm{d}t = 1/2$, and by condition A1, there exists a constant T such that $\sup_{ij} |f'_{ij}(\zeta(t))| \le T$. Therefore,

$$\left| \mathbb{E}\left(\tilde{S}_{ij} - S_{ij} \right) \right| \le h^2 \int_{-\infty}^{+\infty} |t|\Phi(-|t|) \left| f'_{ij}(\zeta(t)) \right| \mathrm{d}t \le Th^2/2.$$

Under conditions A3 and A4, as $N \to \infty$

$$\left\| N^{-1/2}\mathbb{E}\left\{ \widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta}) \right\} \right\| \le N^{-1/2} \sup_{i,j} \left| \boldsymbol{g_{ij}} \right| \sum_{i=1}^{N} Th^2/2 = o(1).$$

In addition, $N^{-1}\mathrm{Var}\{\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta})\} = N^{-1} \sum_{i=1}^{N} \mathrm{Var}\{\sum_{j=1}^{n_i} \boldsymbol{g_{ij}}\, \mathrm{sgn}(-d_{ij})\Phi(-|d_{ij}|)\}$. By Cauchy-Schwartz inequality,

$$\mathrm{Var}\left\{ \widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta}) \right\} \le \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{g_{ij}} \boldsymbol{g_{ij}}^{\mathrm{T}} \mathrm{Var}\left(\tilde{S}_{ij} - S_{ij} \right)$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{j' \ne j}^{n_i} \boldsymbol{g_{ij}} \boldsymbol{g_{ij'}}^{\mathrm{T}} \sqrt{\mathrm{Var}\left(\tilde{S}_{ij} - S_{ij} \right) \mathrm{var}\left(\tilde{S}_{ij'} - S_{ij'} \right)}.$$

For each $j = 1, \ldots, n_i$,

$$\mathrm{Var}\left(\tilde{S}_{ij} - S_{ij} \right) \le \mathbb{E}\left(\tilde{S}_{ij} - S_{ij} \right)^2 = \int_{-\infty}^{+\infty} \left\{ \mathrm{sgn}\left(-d_{ij}\right) \Phi\left(-\left|d_{ij}\right|\right) \right\}^2 f_{ij}(\epsilon)\, \mathrm{d}\epsilon$$

$$= h \int_{-\infty}^{+\infty} \Phi^2(-|t|) f_{ij}\,(ht)\, \mathrm{d}t$$

$$= h \int_{|t| > \Delta} \Phi^2(-|t|) f_{ij}\,(ht)\, \mathrm{d}t + h \int_{|t| \le \Delta} \Phi^2(-|t|) f_{ij}\,(ht)\, \mathrm{d}t$$

$$\le \Phi^2(-\Delta) + h\Delta f_{ij}(\zeta),$$

where Δ is a positive value, and ζ lies between $(-h\Delta, h\Delta)$. Let $\Delta = N^{1/3}$. Under condition A4, $h = o(N^{-1/4})$, then as $N \to \infty$, both $\Phi^2(-\Delta)$ and $h\Delta f_{ij}(\zeta)$ go to 0. By conditions A2 and A3, it is easy to obtain $N^{-1}\mathrm{Var}\{\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta})\} = o(1)$. Therefore, we have $N^{-1/2}\{\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - \boldsymbol{W}(\boldsymbol{\beta})\} \to 0$ as $N \to +\infty$ for any $\boldsymbol{\beta}$. ■

Proof of Theorem 2.3: According to Lemmas 2.1 and 2.2, we have $N^{-1/2}\widetilde{\boldsymbol{W}}(\boldsymbol{\beta_0}) \to N(\boldsymbol{0}, \boldsymbol{\Xi})$, and by the triangle inequality, we obtain that

$$\sup_{||\boldsymbol{\beta} - \boldsymbol{\beta_0}|| \le KN^{-1/2}} ||N^{-1/2}\widetilde{\boldsymbol{W}}(\boldsymbol{\beta}) - N^{-1/2}\widetilde{\boldsymbol{W}}(\boldsymbol{\beta_0}) - N^{1/2}\boldsymbol{A}(\boldsymbol{\beta} - \boldsymbol{\beta_0})|| = o_p(1). \tag{A3}$$

The above two results show that our smoothed WQEEs satisfy the C.1 condition in Ref. [19]. The C.2 condition is for the penalty function $q_\lambda(\cdot)$, which states that:

(a) For non-zero fixed θ, $\lim_{N\to\infty} N^{1/2} q_\lambda(|\theta|) = 0$ and $\lim_{N\to\infty} q'_\lambda(|\theta|) = 0$.
(b) For any $K > 0$, $\lim_{N\to\infty} \sqrt{N} \inf_{|\theta| \le KN^{-1/2}} q_\lambda(|\theta|) \to \infty$.

In our work, we choose the SCAD penalty, that is,

$$q_\lambda(|\theta|) = \lambda I(|\theta| \le \lambda) + \frac{(a\lambda - |\theta|)_+}{a-1} I(|\theta| > \lambda), \quad a > 2.$$

It is easy to see that under condition A5, the condition C.2 holds because $\sqrt{N} q_\lambda(|\theta|) = q'_\lambda(|\theta|) = 0$ for $\theta \neq 0$ and $\sqrt{N} \inf_{|\theta| \le KN^{-1/2}} q_\lambda(|\theta|) = \sqrt{N}\lambda$.

Under conditions A1–A5, we can show that the smoothed WQEE satisfies the conditions C.1 and C.2. Then Theorem 2.3 can be proved by using the same method used in the proof of Theorem 1 in Ref. [19], we omit the details for saving space. ■