



Article Double Penalized Expectile Regression for Linear Mixed Effects Model

Sihan Gao¹, Jiaqing Chen^{1,*}, Zihao Yuan¹, Jie Liu¹ and Yangxin Huang²

- ¹ Department of Statistics, College of Science, Wuhan University of Technology, Wuhan 430070, China; 305856@whut.edu.cn (S.G.); zihaoyuan@whut.edu.cn (Z.Y.); 238645@whut.edu.cn (J.L.)
- ² Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, USA; yhuang@health.usf.edu
- * Correspondence: jqchenwhut@163.com

Abstract: This paper constructs the double penalized expectile regression for linear mixed effects model, which can estimate coefficient and choose variable for random and fixed effects simultaneously. The method based on the linear mixed effects model by cojoining double penalized expectile regression. For this model, this paper proposes the iterative Lasso expectile regression algorithm to solve the parameter for this mode, and the Schwarz Information Criterion (SIC) and Generalized Approximate Cross-Validation Criterion (GACV) are used to choose the penalty parameters. Additionally, it establishes the asymptotic normality of the expectile regression coefficient estimators that are suggested. Though simulation studies, we examine the effects of coefficient estimation and the variable selection at varying expectile levels under various conditions, including different signal-to-noise ratios, random effects, and the sparsity of the model. In this work, founding that the proposed method is robust to various error distributions at every expectile levels, and is superior to the double penalized quantile regression method in the robustness of excluding inactive variables. The suggested method may still accurately exclude inactive variables and select important variables with a high probability for high-dimensional data. The usefulness of doubly penalized expectile regression in applications is illustrated through a case study using CD4 cell real data.

Keywords: double penalized; mixed effects; expectile regression; Lasso algorithm

1. Introduction

Linear mixed effect model (LME) is an important statistical model, which is widely used in various fields. The LME model includes the fixed effect and random effect. The fixed effect is used to represent the general characteristics of the sample population, and the random effect is used to depict the divergences between individuals and the correlation between multiple observations. The structure and property of this model reveal substantial differences compared to the general linear models and complete random coefficient linear models. Compared with other linear models, the inclusion of random effects in the mixed effects model captures the correlation between the observed variables.

Maximum likelihood and least square method are the classical methods of estimation used for the LME model. However, the least square method leads to biased estimates when the data with heavy-tailed distribution or significant heteroscedasticity. Koenker and Bassett (1978) [1] considered the quantile regression to solve such problems by regressing covariates according to the conditional quantiles of response, and capture the regression models under the all quantiles. However, the sparsity of sample variable is an issue which can not be ignored when it involves the correction analysis of variables, that is not all variables have predictive roles on regression analysis. In practical applications, massive candidate variables can be used for modeling analysis and prediction. Retaining the incorrelated variables in the model for prediction is undesirable, where the retention of irrelevant variables will produce the large deviation and non-interpretability of the model.



Citation: Gao, S.; Chen, J.; Yuan, Z.; Liu, J.; Huang, Y. Double Penalized Expectile Regression for Linear Mixed Effects Model. *Symmetry* **2022**, *14*, 1538. https://doi.org/10.3390/ sym14081538

Academic Editors: OPhir Nave and Svetlana Bunimovich-Mendrazitsky

Received: 30 June 2022 Accepted: 25 July 2022 Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). How to select variables effectively is a challenging topic for the linear mixed effects models. Similar to the general linear regression, replacing the different types of penalty terms in the quantile regression can achieve synchrosqueezing for coefficients to reach target of variable selection. For example, Tibshirani (1996) [2] proposed a Lasso method with penalty terms, and selected variables by squeeze parameters sparsely in quantile regression. Quantile Lasso regression constraints some coefficients of irrelevant variables to 0 when the sum of absolute values of coefficients is less than a pre-specified constant, which not only can acquire a more simplified model and select variables simultaneously, also solve the problem of data with sparsity. Next, Zou (2006) [3] gave the proof of the oracle properties of adaptive Lasso in generalized linear models. Biswas and Das (2021) [4] proposed a Bayesian approach of estimating the quantiles of multivariate longitudinal data. Koenker (2004) [5] proposed the L_1 penalty quantile regression based on random effects, which can estimates parameter through weighting random effects information of multiple quantiles. Wang (2018) [6] proposed a new semiparametric approach that uses copula to account for intra-subject dependence and approximates the marginal distributions of longitudinal measurements, given covariates, through regression of quantiles. Wu and Liu (2009) [7] proposed the SCAD and adaptive Lasso penalized quantile regression method, and gave the oracle properties in the situation of variable number unchanged. For different types of data, Peng (2021) [8] illustrated the practical utility of quantile regression in actual survival data analysis through two cases. Li (2020) [9] constructed a double penalized quantile regression method, and used the Lasso algorithm to solve the parameters. It proved that the estimated accuracy of the double penalized quantile regression is better than the other quantile regression. And according to the article above, we find that double penalized quantile regression model lacks accuracy and stability in excluding inactive variables. And expectile regression can accurately reflect the tail characteristics of the distribution, the variables of the model can be accurately selected. Therefore, in order to excluding inactive variables more accurately, this paper cojoin the double penalty terms into the expectile regression to obtain the double penalized expectile regression model.

Newey and Powell (1987) [10] replaced the L_1 -norm loss function to L_2 -norm loss function for weighted least-squares, and proposed expectile regression. It regresses the covariates based on the conditional expectile of the response to obtain the regression model under all expectiles level. Almanjahie et al. (2022) [11] investigated the nonparametric estimation of the expectile regression model for strongly mixed function time series data. Gu (2016) [12] proposed the regularized expectile regression to analysis heteroscedasticity of high-dimensional data. Farooq and Steinwart (2017) [13] analysed a support vector machine type approach for estimating conditional expectiles. Expectile and quantile are metric indicators that capture the tail behavior of data distribution. When covariates make different impacts on the distributions of different response, such as right or left skew, the two metrics can not only decrease the effects of outliers on statistical inference, but also provide a more comprehensive characteristics of entire distribution. Therefore, quantile and expectile regression provide a more comprehensive relation between the covariates and response.

Quantile regression is the generalization of median regression, and expectile regression is the generalization of mean regression, so expectile regression inherits the computational convenience and sensitivity to the observation values. Especially in the financial field, researchers need the sensity of expectile regression to data. For panel data, the model proposed by Schulze and Kauermann (2017) [14] allows multiple covariates, and a semi-parametric approach with penalized splines is pursued to fit smooth expectile curves. For cross-sectional data, expectile regression models and its applications have been studied. Sobotha et al. (2013) [15] discussed the expectile confidence interval based on large sample properties; Zhao and Zhao (2018) [16] proposed the penalized expectile regression model with SCAD penalty with the proof of asymptotic property; Liao et al. (2019) [17] proposed penalized expectile regression with adaptive Lasso and SCAD penalty for variable selection, and gave the proof of oracle properties under independent but different distributions

of error terms. Waldmannetal et al. (2017) [18] proposed a combined Bayesian method with weighted least squares to estimate complex expectile regression. Next, the newly proposed iterative Lasso-expectile regression algorithm is used to solve the estimation of parameters and variables selections. Xu and Ding (2021) [19] combined the elastic network punishment with expectile regression and constructed the elastic network penalized expectile regression model. For the underlying optimization problem, Farooq and Steinwart (2017) [20] proposed an efficient sequential-minimal-optimization-based solver and derived its convergence. For the model selection problem, Spiegel et al. (2017) [21] introduced several approaches on selection criteria and shrinkage methods to perform model selection in semiparametric expectile regression. Expectile regression has also received attention in the economic and financial sector, particularly in actuarial and financial risk management. Daouia et al. (2020) [22] derive joint weighted Gaussian approximations of the tail empirical expectile and quantile processes under the challenging model of heavy-tailed distributions. Ziegel (2013) [23] applied expectile regression to the field of risk measurement.

Expectile regression is widely used in various fields. Including economic field [11,14,17,22,23], biomedical field [12,19] and health field [15,18,21] etc. Therefore, the study of expectile regression has many practical significances, and it is very necessary to study it.

To select the important variables into the model and the inactive variables excluded from the model more accurately, this paper combines the linear mixed effect model with the expectile regression model, and cojoin penalty terms into the estimation of random and fixed effects to construct the double penalized expectile regression for linear mixed effects model. And the iterative Lasso-Expectile regression algorithm is used to solve parameters. The asymptotic property of double penalized expectile regression estimation is proved. The simulation studies will analysis the results of coefficient estimation and variable selection of the method proposed in this paper under different conditions, and the robustness of this method in excluding inactive variables is mainly studied. Finally, based on the research on the real data of CD4 cells, the practical utility difference between the double penalized quantile regression and the double penalized expectile regression are compared.

The rest of this paper is organized as follows. We propose the double penalized expectile regression method and the iterative Lasso expectile regression algorithm in Section 2. The convergence of the algorithm and the asymptotic properties of the model are given in Section 3. In Section 4, we present the simulation studies. And a real data example is illustrated in Section 5. Moreover, this method is compared with the existing double penalized quantile regression method in parameter estimation and variable selection in simulation studies and real data analysis. In Section 6, we give the conclusions. In Appendices, we show the proofs of lemmas and asymptotic properties, and some graphs and tables obtained by simulation studies.

2. Methodologies

In order to get the specific formula of the linear mixed effect double penalized expectile regression model, we introduce the LME model and summarize the estimation methods. And give the specific steps of iterative lasso-expectile regression algorithm. Then we discuss the selection criteria of penalty parameters.

2.1. Model and Estimation

Firstly, we consider the LME model

$$y_{ij} = \mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_{i} + r_{ij}, \ i = 1, 2, \cdots, n, j = 1, 2, \cdots, n_{i}, \sum_{i} n_{i} = N,$$
 (1)

where $\boldsymbol{\beta}(t) = \{\beta_1(t), \dots, \beta_h(t)\}^T$ is the $h \times 1$ vector of fixed effects regression coefficients, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{il})^T$ is a $l \times 1$ vector of random effects. \boldsymbol{x}_{ij}^T is the row vector of the known design matrix, \boldsymbol{z}_{ij} is the $l \times 1$ covariate associated with random effects, and y_{ij} is the *j*th scalar of the *i*th subject's continuous random variable. And we let $\alpha_i \sim N(\mathbf{0}, \mathbf{P})$, $r_i \sim N(\mathbf{0}, \mathbf{Q}_i)$. The Equation (1) can be expressed as

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{r}_i, \ \boldsymbol{\alpha}_i \sim N(\boldsymbol{0}, \boldsymbol{P}), \ \boldsymbol{r}_i \sim N(\boldsymbol{0}, \boldsymbol{Q}_i), \ i = 1, 2, \dots, n$$
(2)

where $X_i = (x_{i1}, x_{i2}, \cdots, x_{in_i})^T$, $Z_i = (z_{i1}, z_{i2}, \cdots, z_{in_i})^T$, $y_i = (y_{i1}, y_{i2}, \cdots, y_{in_i})^T$, and $r_i = (r_{i1}, r_{i2}, \cdots, r_{in_i})^T$.

Next, we let $\mathbf{Z} = \operatorname{diag}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n), \mathbf{r} = (\mathbf{r}_1^{\mathsf{T}}, \mathbf{r}_2^{\mathsf{T}}, \dots, \mathbf{r}_n^{\mathsf{T}})^{\mathsf{T}}, \mathbf{X} = (\mathbf{X}_1^{\mathsf{T}}, \mathbf{X}_2^{\mathsf{T}}, \dots, \mathbf{X}_n^{\mathsf{T}})^{\mathsf{T}},$ and $\mathbf{y} = (\mathbf{y}_1^{\mathsf{T}}, \mathbf{y}_2^{\mathsf{T}}, \dots, \mathbf{y}_n^{\mathsf{T}})^{\mathsf{T}}, \quad \mathbf{\alpha} = (\mathbf{\alpha}_1^{\mathsf{T}}, \mathbf{\alpha}_2^{\mathsf{T}}, \dots, \mathbf{\alpha}_n^{\mathsf{T}})^{\mathsf{T}}, \quad \widetilde{\mathbf{P}} = \operatorname{diag}(\mathbf{P}, \mathbf{P}, \dots, \mathbf{P}),$ $\mathbf{Q} = \operatorname{diag}(\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_n).$ So Considering the matrix form, the LME model (2) can be expressed as

$$y = X\beta + Z\alpha + r, \ \alpha \sim N(\mathbf{0}, \mathbf{P}), \ r \sim N(\mathbf{0}, \mathbf{Q})$$
(3)

Using the maximum likelihood and generalized least square method, the parameters β and α_i can be calculated. According to the known P and Q_i , $i = 1, 2, \dots, n$, consider following joint density function of y_i and α_i :

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}_{i} | \boldsymbol{y}) = \sum_{i=1}^{n} \{ (\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \boldsymbol{\beta} - \boldsymbol{Z}_{i} \boldsymbol{\alpha}_{i})^{\mathrm{T}} \boldsymbol{Q}_{i}^{-1} (\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \boldsymbol{\beta} - \boldsymbol{Z}_{i} \boldsymbol{\alpha}_{i}) + \boldsymbol{\alpha}_{i}^{\mathrm{T}} \boldsymbol{P}^{-1} \boldsymbol{\alpha}_{i} + \log |\boldsymbol{P}| + \log |\boldsymbol{Q}_{i}| \}$$

$$(4)$$

It is possible to determine the β and α_i by minimizing the twice negative logarithm of Equation (4). Equation (4) is not a typical log-likelihood function since α_i are vectors of random effects parameters. To be more specific, the first part in Equation (4) is a weighted residual that accounts for within-subject variation, while $\alpha_i^T P^{-1} \alpha_i$ is a representation of a penalty term resulting from random effects α_i that accounts for between-subject variation.

Minimizing Equation (4) is identical to solving the following mixed model equation [24,25] for given the Q_i and P.

$$\begin{pmatrix} X^{\mathrm{T}}Q^{-1}X & X^{\mathrm{T}}Q^{-1}Z \\ Z^{\mathrm{T}}Q^{-1}X & Z^{\mathrm{T}}Q^{-1}Z + \widetilde{P}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} X^{\mathrm{T}}Q^{-1}y \\ Z^{\mathrm{T}}Q^{-1}y \end{pmatrix}$$

Expressed in matrix form as:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{W}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{W}^{-1} \boldsymbol{y}$$
(5)

$$\hat{\boldsymbol{\alpha}}_i = \boldsymbol{P} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}})$$
(6)

where $W_i = Z_i P Z_i^T + Q_i$, $W = \text{diag}(W_1, W_2, ..., W_n)$. Since this outcome had already been achieved, Robinson [26] credited Henderson [27] with the aforementioned normal equations. The "best linear unbiased predictor" (BLUP), which Goldberger [28] introduced, is used to refer to the random effect estimated value $\hat{\beta}$ and associated estimated value $\hat{\alpha}_i$. And Rao [24] put up a way to demonstrate the following Proposition 1:

Proposition 1. $\hat{\beta}$ solves $\min_{\alpha,\beta} \|y - X\beta - Z\alpha\|_{Q^{-1}}^2 + \|\alpha\|_{\tilde{P}^{-1}}^2$, where $\|x\|_T^2 = x^T Tx$.

Although the implicit estimation solution of random effects may be distinct from others, seeing the random effects estimation value as a penalized least squares estimator offers useful insights for the addition of punishment items. By shrinking the unrestricted $\hat{\alpha}$ reaches the expected value, and both accuracy of the estimator of $\hat{\beta}$ and the individual fixed effect estimation will be improved. According to the classical statistical conclusion, in the case of non-informative priors, the posterior expectations of parameters of Bayesian estimation of LME are also solutions (5) and (6).

2.2. Double Penalized Expectile Regression

Li (2020) [9] proposed a double penalized quantile regression estimation for the LME model that finds $\hat{\alpha}_i \in \mathbb{R}^l$ and $\hat{\beta} \in \mathbb{R}^h$, i = 1, ..., n that minimize

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \gamma_{\theta}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_i) + \lambda_{\beta} \sum_{k=1}^{h} |\beta_k| + \lambda_{\alpha} \sum_{i=1}^{n} \sum_{\mathrm{T}=1}^{l} |\alpha_{it}|$$
(7)

where the so-called check function of under θ th level quantile regression is denoted by $\gamma_{\theta}(k) = k(\theta - I(k < \mathbf{0}))$. Equation (7) can simultaneously perform a variable selection operation and estimate the mixed expectile functions of the response variable, as stated in Li (2020) [9].

Considering the lack of robustness of quantile regression for excluding inactive variables. Inspired by Newey and Powell (1987) [10], we can use the following method for the conditional expectile functions of the response of the *j*th observation on the *i*th individual y_{ij} ,

$$Q_{y_{ij}}(\tau | \mathbf{x}_{ij}, \mathbf{\alpha}_i) = \mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathrm{T}} \mathbf{\alpha}_i$$
(8)

The same individual is connected to each component of $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{il})^{\mathrm{T}}$. We suggest a double penalized expectile regression method for the LME model based on Equation (8), which find $\hat{\boldsymbol{\alpha}}_i \in \mathbb{R}^l$ and $\hat{\boldsymbol{\beta}} \in \mathbb{R}^h$, $i = 1, \dots, n$ that minimize

$$L(\boldsymbol{\beta},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho_{\tau}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i) + \lambda_{\beta} \sum_{k=1}^{h} |\boldsymbol{\beta}_k| + \lambda_{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{l} |\boldsymbol{\alpha}_{it}|$$
(9)

where $\rho_{\tau}(u) = |\tau - I(u < 0)|u^2$ signifies the check function, and $\tau \in (0, 1)$.

Compared with Equation (7), the expectile regression has the sensitivity to extreme values, and the square loss function used in the regression has the computational advantage. When discussing the asymptotic properties of the estimation, the covariance matrix of the asymptotic distribution does not need to calculate the density function of the residual. Therefore, compared with quantile regression, expectile regression depends more on the overall distribution, and similar to Equation (7), our method also considers selecting variables when estimating coefficients.

2.3. Iterative Lasso-Expectile Regression Algorithm

Obviously, it is very difficult to directly estimate parameters β , α_i , λ_β and λ_α of the iterative double penalized expectile regression model. We propose the Lasso-expectile regression algorithm to solve this optimization problem by selecting one variable and fixing another to solve α_i and β , it is equivalent to solving the general lasso expectile model. The iterative series can not guarantee the convergence to the global optimum, the objective function increases monotonically and achieves the maximum value. Therefore, the algorithm terminates after limited iterations and reaches a local optimum (the detailed proof in Section 3).

It is more efficient to combine the iterative approach with the adjustment of λ_{β} than to solve $\hat{\beta}$ for a fixed λ_{β} since this algorithm can find the solution paths of $\hat{\beta}$ with λ_{β} when provided α . According to several selection criteria, we choose the best λ_{β} to fit β and the best λ_{α} to fit α_i . The SIC [29] (Schwarz Information Criterion) and GACV [30] (generalized approximate cross-validation criterion) are the two criteria that are most frequently employed for expectile regression, respectively.

$$\operatorname{SIC}(\lambda_{\beta}) = \ln(S_K/N) + \frac{\ln(N)}{2N}|K|$$
(10)

$$GACV(\lambda_{\beta}) = \frac{S_M}{N - |M|}$$
(11)

where $S_K = \sum_{i=1}^n \sum_{j=1}^{n_i} \gamma_{\theta}(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta} - \mathbf{z}_{ij}^T \hat{\mathbf{\alpha}}_i), N = \sum_{i=1}^n n_i$, and the dimension of the process *K* is |K|, which is the same as one of the non-zero components of $\boldsymbol{\beta}$. Only when the result of SIC or GACV is at the inflection point is the ideal parameter obtained.

In fact, we can present a different set of tuning parameters firstly, such as dividing (0, a] evenly into *b* parts, and then select the best tuning parameter based on these two criteria. In order to solve the parameter β for the variables $\alpha_i = \hat{\alpha}_i$, $i = 1, 2, \dots, n$, we compared these two criteria. The solution to α_i might also be derived similarly provided $\beta = \hat{\beta}$.

We summarize the above iterative algorithm in the following Algorithm 1 steps. The specific steps of iterative Lasso-expectile regression algorithm are as follows.

Algorithm 1. Iterative Lasso-Expectile Regression Algorithm.
Input: y, X, Z
Output: $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \lambda_{\beta}, \lambda_{\alpha}$.
Step 1: Give the initial value, $\hat{\pmb{lpha}}_i^{(0)}=0, i=1,\cdots,n$, and get the standard Lasso solution $\hat{\pmb{eta}}^{(0)}$
according to $\hat{\beta}^{(0)} = \operatorname{argmin} L^*(\boldsymbol{\beta}, 0);$
Step 2: Iterate the following two Lasso optimization steps, $v = 0, 1, \cdots$,
• $\hat{\alpha}^{(v+1)} = \operatorname{argmin} L^*(\hat{\beta}^{(v)}, \alpha)$, and the modified residual is $f^{(v)} = y_{ij} - x_{ij}^T \hat{\beta}^{(v)}$;
• $\hat{oldsymbol{eta}}^{(v+1)} = \operatorname{argmin} L^*(oldsymbol{eta}, \hat{oldsymbol{a}}^{(v+1)})$, and get a new response variable $y_{ij}^* = y_{ij}^* - z_{ij}^T \hat{oldsymbol{a}}_i^{(v+1)}$;
Step 3: Terminate when $\sum_{k=1}^{h} \left \left(\hat{\beta}_{k}^{(v+1)} - \hat{\beta}_{k}^{(v)} \right) \right / h < \delta$ for a pre-specified small value δ .

Then we use an example to show the solution path in each iteration step for SIC and GACV. We consider the following the process of model data generating:

$$y_{ij} = \mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_i + r_{ij}$$

where $\boldsymbol{\beta} = (3, -1.5, 0, 0, 2, 0, 0, 0)^{\mathrm{T}}$, $\boldsymbol{z}_{ij}^{\mathrm{T}} = (1, x_{ij1})$, $\boldsymbol{x}_{ij}^{\mathrm{T}} = (x_{ij1}, x_{ij2}, \dots, x_{ij8})$, $i = 1, 2, \dots, 10$, $j = 1, 2, \dots, 20$ are iid from standard normal distribution, and the random effects are iid from $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1})^{\mathrm{T}} \sim N_2(0, I)$, $r_{ij} \sim N(0, 4)$. We use double penalized Lasso ecpectile regression (DLER) and double penalized Lasso quantile regression (DLQR) methods to study the two criteria, and get the following results.

Figures A1 and A2 (in the Appendix B) show the search paths of the DLER method for penalty parameters α and β under the SIC and GACV criteria in the first iteration respectively. It can be seen in the graph that the penalty parameter search paths of SIC and GACV are obviously different, but the corresponding optimal penalty parameter values are close.

Figures A3 and A4 (in the Appendix B) show the results of DLQR method in the first iteration process respectively. The loss functions of DLER method and DLQR method are obviously different in nature. Under different penalty parameters, the criterion values of the two methods are not comparable, so they cannot be compared and analyzed horizontally from the numerical perspective of the two criteria. From the path of fixed effect coefficient, it can be seen that with the increase of penalty coefficient λ , DLER method will quickly compress the inactive covariate coefficients to 0, while the coefficients of important covariates change slowly around the true value. However, The DLQR method does not fluctuate much with changes in the penalty coefficient. It shows that the DLER method has an absolute advantage over DLQR in terms of variable selection.

3. Asymptotic Properties

In this part, we give the convergence of iterative Lasso expectile regression algorithm and the asymptotic property of double penalized expectile regression.

3.1. Convergence of Iterative Lasso-Expectile Algorithm

In order to prove the asymptotic properties of the double penalized expectile regression method. The following lemmas are provided:

Lemma 1. There exists a single $\beta(\alpha)$ that minimizes the cost function $L(\beta, \alpha)$ for any given λ_{β} , λ_{α} , if $L(\beta, \alpha)$ is a continuous and strictly convex function of α and β . The same conclusion also applies to α .

The Proof of Lemma 1 is given in Appendix A.

Lemma 2. For a given λ_{β} , λ_{α} , definition $\eta : \mathbb{R}^d \to \mathbb{R}^d$ is the mapping of the above iterative algorithm for updating process of $(\beta, \alpha) \to (\beta^{(1)}, \alpha^{(1)})$ in one step, then η is continuous.

The Proof of Lemma 2 is given in Appendix A.

Lemma 3. The sequence solution $(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\alpha}}^{(r)})$ provided in the iterative Lasso expectile algorithm reduces the objective function $(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\alpha}}^{(r)})$ and $L(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\alpha}}^{(r)})$ converges to $(\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\alpha}})$ under the assumption that there is a unique $(\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\alpha}})$ for the given λ_{β} and λ_{α} to minimize $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

The Proof of Lemma 3 is given in Appendix A.

3.2. Asymptotic Properties of DLER

In order to obtain the asymptotic properties of the double penalized expectile regression method, we first propose the following regularization assumptions:

(A1) y_{ij} has independent conditional distribution function $F_{ij}(\cdot)$ and density function $f_{ij}(\cdot)$, given $\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{\alpha}_i$, and $0 < f_{ij}(\mathbf{\kappa}_{ij}(\tau)) < +\infty, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m$.

Where $\boldsymbol{\kappa}_{ij}(\tau) = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}(\tau) + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_{i}, \quad \mathbb{V}\mathrm{ar}[\boldsymbol{\Pi}_{\tau}(r_{i\tau})r_{i\tau}] = \mathbb{E}[\boldsymbol{\Pi}_{\tau}(r_{i\tau})r_{i\tau}r_{i\tau}^{\mathrm{T}}\boldsymbol{\Pi}_{\tau}(r_{i\tau})] = \boldsymbol{\Lambda}_{i\tau},$ $r_{i\tau} = (r_{i1\tau}, \cdots, r_{im\tau})^{\mathrm{T}}, \quad r_{ij\tau} = y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_{\tau}, \quad \boldsymbol{\Pi}_{\tau}(r_{i\tau}) = [diag(\varphi_{\tau}(r_{ij\tau}))]_{j=1}^{m};$

(A2) Assuming that α_{it} , $i = 1, 2, \dots, n$, $t = 1, 2, \dots, l$ are independent, and their distribution function and density function are $G_{it}(\cdot)$ and $g_{it}(\cdot)$ respectively, $0 < g_{it}(0) < +\infty$;

(A3) Definition $\vartheta = \tau(1 - \tau)$, $\Delta = diag(f_{ij}(\kappa_{ij}(\tau)))$, and there are the positive definite matrixes

$$P_{0}(\tau) = \lim_{\substack{m \to \infty \\ n \to \infty}} \vartheta m^{-1} \begin{pmatrix} \mathbf{Z}^{\mathsf{T}} \boldsymbol{\Lambda}_{\tau} \mathbf{Z} & \mathbf{Z}^{\mathsf{T}} \boldsymbol{\Lambda}_{\tau} \mathbf{X} / \sqrt{n} \\ \mathbf{X}^{\mathsf{T}} \boldsymbol{\Lambda}_{\tau} \mathbf{Z} / \sqrt{n} & \mathbf{X}^{\mathsf{T}} \boldsymbol{\Lambda}_{\tau} \mathbf{X} / n \end{pmatrix}$$

$$\mathbf{P}_{1}(\tau) = \lim_{\substack{m \to \infty \\ n \to \infty}} m^{-1} \begin{pmatrix} \mathbf{Z}^{\mathrm{T}} \mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})] \mathbf{Z} & \mathbf{Z}^{\mathrm{T}} \mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})] \mathbf{X}/\sqrt{n} \\ \mathbf{X}^{\mathrm{T}} \mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})] \mathbf{Z}/\sqrt{n} & \mathbf{X}^{\mathrm{T}} \mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})] \mathbf{X}/n \end{pmatrix}$$

where $\Lambda_{\tau} = \mathbb{V}ar[\Pi_{\tau}(r_{\tau})r_{\tau}] = diag[\Lambda_{i\tau}]_{i=1}^{n}$;

(A4) $\max_{1 \le i \le n, 1 \le j \le m} ||x_{ij}|| / \sqrt{nm} \to 0$, $\max_{1 \le i \le n, 1 \le j \le m} ||z_{ij}|| / \sqrt{m} \to 0$.

The above assumptions (A1) and (A3) are the standard conditions of the panel data models, where assumption (A1) is common in the literature of quantile regression, which not only ensures the independence between observation individuals, but also allows the heterogeneity within individuals; (A3) gives the full rank condition, according to Lindeberg-Feller central limit theorem, when $\tau = 0.5$, it will be simplified by $P_1(\tau)$, and then (A3) will be simplified.

We give the asymptotic properties of the double penalty expectile regression estimation of Equation (7), and consider the following objective function:

$$\Gamma_{nm}(\boldsymbol{u}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\rho_{\tau}(y_{ij} - \kappa_{ij}(\tau) - \frac{x_{ij}^{\mathsf{T}}\boldsymbol{u}^{(1)}}{\sqrt{nm}} - \frac{z_{ij}^{\mathsf{T}}\boldsymbol{u}^{(2)}}{\sqrt{m}}) - \rho_{\tau}(y_{ij} - \kappa_{ij}(\tau)) \right] \\ + \lambda_{\beta} \sum_{k=1}^{h} \left(\left| \beta_{k} + \frac{u_{k}^{(1)}}{\sqrt{nm}} \right| - \left| \beta_{k} \right| \right) + \lambda_{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{l} \left(\left| \alpha_{it} + \frac{u_{it}^{(2)}}{\sqrt{m}} \right| - \left| \alpha_{it} \right| \right)$$
(12)

where $\Gamma_{nm}(u)$ is a convex function whose minimum value is

$$\hat{\boldsymbol{u}} = \begin{pmatrix} \hat{\boldsymbol{u}}^{(1)} \\ \hat{\boldsymbol{u}}^{(2)} \end{pmatrix} = \begin{pmatrix} \sqrt{nm}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \\ \sqrt{m}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \end{pmatrix}$$
(13)

Theorem 1. Under assumptions (A1)–(A4), when, $n \to \infty$, $m \to \infty$, if, $\lambda_{\beta}/\sqrt{mn} \to \lambda_1 \ge 0$, $\lambda_{\alpha}/\sqrt{m} \to \lambda_2 \ge 0$, and there is the a > 0, and $n^a/m \to 0$ is satisfied, so we have

$$\hat{\boldsymbol{u}} = \operatorname{argmin}\Gamma_{mn}(\boldsymbol{u}) \stackrel{a}{\rightarrow} \operatorname{argmin}\Gamma_{0}(\boldsymbol{u})$$

where

$$\Gamma_{0}(\boldsymbol{u}) = -2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{W} + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}_{1}\boldsymbol{u} + \lambda_{1}\sum_{k=1}^{h} \left[u_{k}^{(1)}\mathrm{sgn}(\beta_{k})I(\beta_{k}\neq 0) + \left| u_{k}^{(1)} \right| I(\beta_{k}=0) \right] + \lambda_{2}S^{\mathrm{T}}\boldsymbol{u}^{(2)}$$
$$\boldsymbol{W} \sim N(0, \boldsymbol{P}_{0}), S \sim N(0, diag(4G_{it}(0)(1-G_{it}(0))))$$

According to the aforementioned theorem, for non-zero coefficients, the double penalized estimation resulted in bias, and the bias degree was regulated by the tuning parameter. The proof of Theorem 1 is given in the Appendix A.

4. Simulation Study

In this section, we illustrate the performance of the proposed double penalized expectile regression. In order to study the impact of the SIC criteria and the GACV criteria on double penalized Lasso expectile regression model and double penalized Lasso quantile regression model [9]. Simulation 1 is used to compare the two methods with different expectile levels and signal-to-noise ratio. And briefly denoted as DLER-SIC, DLER-GACV, DLQR-SIC, and DLQR-GACV; To illustrate the robustness of the DLER method in excluding inactive variables, simulation 2 and simulation 3 are given to study the impact of random effects on the coefficient estimation, comparing the performance of DLER and DLQR under different random effects. As well as comparing the obtained experimental values at different error distributions and different model sparsity to illustrate the advantage of DLER in variable selection; in order to illustrate the advantage of the DLER method in high-dimensional data, simulation 4 compares the two methods when dimensions of covariate is larger than the sample size.

In order to evaluate the accuracy of model coefficient estimation, we use mean square error (MSE) as the evaluation index. MSE is defined as:

$$MSE = \left(\hat{\boldsymbol{\beta}}^{t} - \boldsymbol{\beta}\right)^{\mathrm{T}} \Sigma^{-1} \left(\hat{\boldsymbol{\beta}}^{t} - \boldsymbol{\beta}\right)$$
(14)

where $\Sigma = (\rho^{|k-h|})_{8\times8}$, and $\hat{\beta}^t$ is the estimator of β in the sth simulation. SD stands for the 100-repetition Bootstrap standard deviation. Corr expresses the average proportion of important variables that were entered into the model correctly, and Incorr expresses the average proportion of inactive variables that were entered into the model

incorrectly. $X_1(SD_1), X_2(SD_2), X_5(SD_5)$ represent the total number of choosing each important variable into the model correctly and its bootstrap standard deviation respectively. $X_3(SD_3), X_4(SD_4), X_6(SD_6), X_7(SD_7), X_8(SD_8)$ represent the total number of correctly excluded each redundant variable out of the model and its bootstrap standard deviation respectively. Int (SD_{int}) represent the total number of choosing intercept into the model and its bootstrap standard deviation.

Simulation 1. The impact of "signal-to-noise ratio" on estimation.

We compared the effectiveness of the two tuning parameter criteria, SIC and GACV. The process for producing the model data listed below was rated as

$$y_{ij} = \beta_0 + \boldsymbol{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_i + \sigma r_{ij}$$
(15)

We let $\mathbf{x}_{ij}^{T} = (x_{ij1}, x_{ij2}, \dots, x_{ij8}), i = 1, 2, \dots, 20, j = 1, 2, \dots, 25,$ $\mathbf{Z}_{ij}^{T} = (x_{ij1}, x_{ij2}, \dots, x_{ij5}), \text{ and } X_1, \dots, X_8 \text{ are generated from } N(0, 1) \text{ with correlation}$ between X_k and X_l being $\rho^{|l-k|}$, $\rho = 0.5$. This paper sets the random effects are iid from $\boldsymbol{\alpha}_{i} = (\alpha_{i0}, \alpha_{i1}, \cdots, \alpha_{i5})^{\mathrm{T}} \sim N_{6}(\mathbf{0}, \mathbf{P}), \text{ where } \mathbf{P} = diag(1, 1, 1, 1, 0, 0). \text{ And } \beta_{0} = 0, \boldsymbol{\beta} = 0$ $(3, -1.5, 0, 0, 2, 0, 0, 0)^{T}$. The error terms are iid from $r_{ij} \sim N(0, 1)$, we set $\tau = 0.25, 0.5, 0.75$, and $\sigma = 1, 2, 3$ to compare the estimation methods. In every simulation, we consider the estimator of coefficients to be 0 if its absolute value is less than 0.1. We set $\varepsilon = 10^{-3}$ in the iterative Lasso algorithm of DLER as the same settings in the iterative algorithm of DLQR of Li (2020) [8]. Tables A1 and A2 (in the Appendix C) give the results of coefficient estimation and variable selection, respectively.

According to Table A1, the estimation accuracy of the model is first analyzed. When σ is fixed, the MSE of the two methods is almost reaches the minimum at $\tau = 0.5$, the estimation accuracy of the two methods is better, while the accuracy at the other fractile is slightly worse, but the difference is not obvious, for example when $\sigma = 1$, MSE of DLER-SIC method is 0.292 at $\tau = 0.5$, while at the other expectile levels are 0.306 and 0.374, respectively. In this case, this shows that the DLER method has the same accuracy as the DLQR method in parameter estimation. In addition, according to the results of the two tunning parameter criterions, the estimation accuracy under the SIC is better than the GACV whether the DLER model or the DLQR model.

Next, analyze the accuracy of variable selection. According to the SD of each variable estimation in Table A1, when fixed σ , with the increase of τ , SD decreases gradually, and the accuracy of variable selection of the two models is gradually increased. For each important variable, according to Table A2, the DLER method can select 99% of the important variables into the model. In addition, for each inactive variable, comparing the results in the table, shows that when τ is fixed, with the increase of σ , the SD is also increasing, indicating that the accuracy of DLQR method is gradually weakened in excluding inactive variables out of the model. For DLER method, the SD is the largest at $\sigma = 2$, DLER method has the low accuracy, it is still better than DLQR at this time, especially at $\sigma = 3$, the SD is the smallest, indicating that with the increase of signal-to-noise ratio, DLER method is better than DLQR method in excluding the redundant variables. Combined with the results of Table A2, the same conclusion can be obtained.

Therefore, for the parameter estimation of the methods, when the signal-to-noise ratio with relatively large, the DLER method is better than the DLQR method. In terms of model variables selection, whether in SIC and GACV criteria, almost 99% of important variables can be selected in both methods. For the ability to exclude the redundant variables, the DLER method is more advantageous than the DLQR method. Therefore, in terms of the accuracy of parameter estimation, DLER method and DLQR method have the same effect. However, the DLER method is better than the DLQR method in terms of estimated stability and excluding inactive variables.

Simulation 2. The influence of random effects.

We use the simulation to show the influence of random effects on DLER-SIC, DLER-GACV, DLQR-SIC and DLQR-GACV. The data generation model is Equation (15) with the fixed $\sigma = 1$. And consider the covariance matrix of the random effect *P* with three forms:

$$P_1 = \text{diag}(1, 1, 1, 1, 0, 0)$$

 $P_2 = \text{diag}(2, 2, 2, 2, 0, 0)$
 $P_3 = \text{diag}(3, 3, 3, 3, 0, 0)$

With the increasing of random effects, we get the results of 100 repetitions simulations with $\tau = 0.25, 0.5, 0.75$. Tables A3 and A4 (in the Appendix C) give the results of estimation and variable selection of different methods under different random effects at $\tau = 0.5$.

According to Table A3, with the increasing interference of random effects, although the estimation accuracy of the two methods is decreasing, and the accuracy of the fixed effect coefficient is decreasing, especially the first two important covariates interfered by random effects. However, in term of variable selection, DLER method has little change in the accuracy of variable selection, and almost all the important variables can be choosing into the model, and the ability to exclude the inactive variables is still better than DLQR method. In particular, the DLER-SIC method, combined with Table A4, shows that the proportion of correctly choosing variables is 99%, and the accuracy of excluding the redundant variables is almost above 90%. In general, since the double penalized expectile regression takes into account the random effects in the accuracy of variables selection, but it will still be affected by random effects to a degree in the accuracy of fixed effect coefficient estimation. Therefore, in terms of the accuracy of parameter estimation, DLER method and DLQR method have the same effect. However, DLER is still more robust than DLQR in variable selection, even if the interference of random effects is added to the model.

Simulation 3. The case of different model sparsity and error distribution.

We compare DLQR and DLER under different model sparsity and error distribution. The data generation model is Equation (15), considering the fixed effect as the following three cases

- (1) Dense $\boldsymbol{\beta} = (1, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$
- (2) Sparse $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^{\mathrm{T}}$
- (3) High Sparse $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0, 0)^{\mathrm{T}}$

We consider $\sigma = 1$, $\tau = 0.5$, P = diag(2, 2, 2, 2, 0, 0), the distribution of error term respectively comes form N(0, 1), t(3) and Cauchy(0, 1). Comparing the models DLER-SIC, DLER-GACV, DLQR-SIC and DLQR-GACV. Tables shows the results of the three models by 100 repetitions simulations. Tables A5 and A6 show the results of coefficient estimation and variable selection under the dense model, Tables A7 and A8 show the results under the sparse model, Tables A9 and A10 show the results under the highly sparse model. (See Appendix C for tables).

According to Table A5. At this time, all variables are important variables. With the change of error distribution, MSE are increasing, and the estimation accuracy of the two methods are decreased. We find that DLER method and DLQR method have the same effect in term of the accuracy of parameter estimation. In addition, considering the accuracy of variable selection of the two methods. Although the two methods cannot completely choose all the important variables into the model, it can be known from Table A6 that the average of the correct variables retained by them is more than 7.6, which is very close to the true value 8. Moreover, when the error term is adjusted from the normal distribution to the heavy-tailed distribution, the change of DLER is the smallest in all methods, so it is weak on the influence of different error distributions. In summary, DLER is robust to the change of error distribution on variable selection.

Next, the results of sparse model and highly sparse model are analyzed. Firstly, the estimation accuracy of DLER method and DLQR method is analyzed. The results are similar to the dense model, with the error distribution becoming more complex, MSE are increasing, and the estimation accuracy of the two methods is decreased. Next, considering the accuracy of the variable selection in model, according to Tables A7 and A8, when the error obeys the normal distribution, for the sparse model, the two methods have little difference in the accuracy of excluding the inactive variables. However, when the error obeys distribution t(3) or Cauchy(0, 1), the DLER method is significantly better than DLQR. Especially in the highly sparse model with error distribution Cauchy(0, 1), DLER excluding the inactive variables advantage is more obvious, especially DLER-SIC method. It shows that expectile regression is quite robust than quantile regression.

Simulation 4. The case of high dimensional data.

High-dimensional data is widely available in the stock exchange market, biomedicine, aerospace and other fields, so the modeling and analysis of high-dimensional data has very important practical significance. Next, we investigate the performance of the proposed model in the high-dimensional scenarios of selecting variables. The data generation model is still the Equation (15), we reduce the sample size to n = 10, m = 10, that is, the total sample size is 100. In addition, 102 independent noise variables X_9 , X_{10} , \cdots , X_{110} are added to the above sparse model, all variables are independent and identically distributed in $N(0, 0.25^2)$, thus the total number of variables is 110, larger than the total sample size, and $\beta_{110\times1} = (3, 1.5, 0, 0, 2, 0, 0, 0, \cdots, 0)^T$. There are three real important covariates and 107 redundant covariates. In addition, we set $\sigma = 0.5$, P = diag(1, 1, 1, 1, 0, 0). Tables A11 and A12 (in the Appendix C) show 100 repeated results of the two methods at $\tau = 0.5, 0.75$, where $X_0(SD_0)$ denote the average and bootstrap standard deviation of all redundant variables being correctly excluded out of the model.

Firstly, analyze the estimation precision of the two methods. According to Table A11, in the situation of fixed quantile, when the dimension of covariates is larger than the sample size, the MSE is larger than the previous simulation, the change of the MSE value of the DLER method is less than DLQR, and the MSE of the DLQR method is significantly larger than the DLER, indicating that although the estimation accuracy of the two methods is decreasing, the DLER method is significantly better than the DLQR method, and the stability of the DLER method is better than the DLQR method. Next, analyze the accuracy of variable selection. According to Tables A11 and A12, the DLER method can ensure that the proportion of excluding redundant variables is more than 95% under three expectile levels. When the expectile level is fixed, DLER method has absolute advantages over DLQR in excluding redundant variables, especially DLER-GACV method can exclude more than 97% of the inactive variables at $\tau = 0.5$. Therefore, when the dimension of covariates is larger than the sample size, DLER method is superior to DLQR in terms of estimation accuracy and excluding the inactive variables both at median $\tau = 0.5$ and extreme expectile level $\tau = 0.75$.

5. Application

CD4 cells play an important role in determining the efficacy of AIDS treatment and the immune function of patients, so excluding inactive variables is important for analyzing CD4 cell data. We applied the model to the real data of CD4 cell count. For a complete description of the data set, please refer to Diggle P.J' s homepage: https://www.lancaster. ac.uk/staff/diggle/, accessed on 30 June 2022, and we select a part of this data set. The response variable is the open-root conversion of CD4 cell count. The variables in the data set include the time of seroconversion (time), the age relative to a starting point (age), the smoking status depicted by the number of packets smoked (smoking), the number of sexual partners (sex partner), and the depression state and depression degree (depression). Choosing time, smoking, age and sex partner as important fixed effect to determine the CD4

cell numbers, where time and age are important random effects. On this basis, consider the following model:

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \mathbf{Z}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_{i} + r_{ij}$$

where Y_{ij} is the *j*th observation of the *i*th individual, X_{ij}^{T} is the observation of explanatory variables, and Z_{ij}^{T} is a subset of X_{ij}^{T} . We set that the threshold of β is 0.05, and the threshold of α_i is 0.1. Table A13 (in the Appendix C) gives the results of the four methods at $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$.

From the results of Table A13, the double penalized expectile regression method proposed can give the results of variable selection while estimating the coefficients. The variable with a value of 0 in the table indicates that it can be excluded from the model. It can be found that both estimation methods excluded variables X_3 and X_4 from the model. In the four variables, only time and smoking will affect CD4 cells. From the sign of coefficient estimation, the value of time is always less than 0, indicating that time has a negative influence on CD4 cells. The values of smoking are all larger than 0, indicating that smoking has a positive influence on CD4 cells.

Analysis of the numerical changes. For variable X_2 , under the SIC criterion, with the changes of expectile level, the DLER method fluctuates less, and the DLER is relatively stable. Indicating that the stability of the DLER-SIC method is slightly stronger. For variable X_1 , whether it is the SIC criteria or the GACV criteria, the numerical fluctuation of the DLER method is less than the DLQR method, indicating that the stability of DLER for estimation of variable X_1 is better than that of DLOR.

This shows that DLER method has strong practical utility for CD4 data in excluding inactive variables and analyzing its influencing factors.

6. Conclusions

In this paper, we propose the double penalized expectile regression method for linear mixed effects model, which imposes penalties on both fixed effects and random effects, fully accounting for the random effects in coefficient variable selection and estimation. The model proposed in this paper is found to be highly robust in excluding inactive variables after simulations and application studies. The conclusion is also supported by comparison with double penalized quantile regression method, through the comparison of the results of the two estimation methods, it can be found that the double penalized expectile regression has the same effect as the double penalized quantile regression in term of the accuracy of parameter estimation, but has absolute advantage in selecting variables, especially in excluding inactive variables.

When the signal-to-noise ratio is different, the precision of coefficient estimation and the accuracy of variable selection will be different. When the signal-to-noise ratio is larger, the proposed method outperforms the quantile model in terms of coefficient estimation ability and excluding inactive variables. In addition, because the double penalized expectile regression select variable for both fixed and mixed effects, it is almost undisturbed to random effects in terms of variable selection accuracy. When comparing the dense and sparse models, it is found that the sparser the model and the more complex the error term distribution, the advantage of the new method is more obvious, indicating that it has strong robustness. When the dimension of covariates is larger than the sample size, the method has obvious advantages in the estimation precision of coefficients and excluding inactive variables. Finally, when analyzing the real data, it is found that the new method has strong practical utility for the longitudinal data of CD4 cells, which can exclude its inactive variables and analyze the influence trend of various factors, therefore obtain accurate medical judgment. **Author Contributions:** Conceptualization, J.C. and S.G.; methodology, S.G.; software, S.G.; validation, Z.Y., J.L. and Y.H.; formal analysis, S.G.; investigation, S.G.; resources, Z.Y.; data curation, J.L.; writing—original draft preparation, S.G.; writing—review and editing, S.G.; visualization, Z.Y.; supervision, J.C.; project administration, Y.H.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant 81671633 to J. Chen.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data in this paper have been presented in the manuscript.

Acknowledgments: Many thanks to reviewers for their positive feedback, valuable comments and constructive suggestions that helped improve the quality of this article. Many thanks to editors' great help and coordination for the publish of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Lemma 1. *First, prove the existence of the conclusion.*

We give $\boldsymbol{\alpha}$ and select $\boldsymbol{\beta}^0 \in \mathbb{R}^h$ randomly. Define

$$M_{\boldsymbol{\beta}} = \left\{ \boldsymbol{\beta} \in \mathbb{R}^{h} \middle| L(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq L(\boldsymbol{\beta}^{0}, \boldsymbol{\alpha}) \right\}$$

The continuity of *L* shows that M_{β} is a closed area. For a given $\lambda_{\beta} > 0$, then

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq L(\boldsymbol{\beta}^0, \boldsymbol{\alpha}) \Rightarrow \|\boldsymbol{\beta}\|_1 \leq L(\boldsymbol{\beta}^0, \boldsymbol{\alpha})/\lambda_{\boldsymbol{\beta}}$$

 M_{β} is contained in the spherical region of k_l , so M_{β} is bounded. Therefor there exists β in M_{β} such that $L^*(\beta, \alpha)$ reaches the minimum.

Then prove the uniqueness of the conclusion. It can be easily obtained from the strictly convex function of β of *L*. Similar to this, there is a specific $\alpha(\beta)$ such that $L(\beta, \alpha)$ reaches the minimum value for a given β and $\lambda_{\alpha} > 0$. \Box

Proof of Lemma 2. We can know that $(\beta^{(1)}, \alpha^{(1)})$ is a composite mapping that only depends on β from the process of iterative algorithm, so all that must be demonstrated is that both $\alpha^{(1)} \rightarrow \beta^{(1)}$ and $\beta \rightarrow \alpha^{(1)}$ are continuous mappings. In addition, since given β to solve $\alpha^{(1)}$, it is symmetrical to make $L(\beta, \alpha^{(1)})$ reach the minimum with given $\alpha^{(1)}$ to solve $\beta^{(1)}$ to make $L(\beta, \alpha^{(1)})$ reach the minimum. Just has to be demonstrated that mapping $\alpha^{(1)} \rightarrow \beta^{(1)}$ continuous.

For simplicity, here we omit the superscript. Next define the function $g(\alpha) = \inf_{\beta} L(\beta, \alpha)$. *g* is a convex continuous function because the function $L(\beta, \alpha)$ is convex.

Next, we prove that $\beta(\alpha_n) \rightarrow \beta(\alpha) = \beta$ is true for every sequence $\alpha_n \rightarrow \alpha$. Since *g* is a continuous function and $\exists N$ for $\forall r > 0$, when $n \ge N$ we have $g(\alpha_n) \le g(\alpha) + r$, that is, $L(\beta(\alpha_n), \alpha_n) \le L(\widetilde{\beta}, \alpha) + r$. And define

$$M = \max\left\{\max_{n=1,2,\cdots,N} L(\boldsymbol{\beta}(\boldsymbol{\alpha}_n), \boldsymbol{\alpha}_n), L(\boldsymbol{\widetilde{\beta}}, \boldsymbol{\alpha}) + r\right\},\$$
$$A = \{(\boldsymbol{\beta}, \boldsymbol{\alpha}) | L(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq M\}$$

Since *A* is a bounded closed set with sequence $(\beta(\alpha_n), \alpha_n)$ for a given λ_{α} , and λ_{β} , there is a convergent subsequence $\beta(\alpha_{n_k})$ whose limit value is $\overline{\beta}$. From the continuity of *L* and *g*, we can get

$$L(\widetilde{\boldsymbol{\beta}},\boldsymbol{\alpha}) = g(\boldsymbol{\alpha}) = \lim_{h \to \infty} g(\boldsymbol{\alpha}_{n_h}) = \lim_{h \to \infty} L(\boldsymbol{\beta}(\boldsymbol{\alpha}_{n_h}), \boldsymbol{\alpha}_{n_h}) = L(\overline{\boldsymbol{\beta}}, \boldsymbol{\alpha})$$

Lemma 1 demonstrates that the point at which $L(\beta, \alpha)$ reaches the minimal value for a given α is unique, and from the equation $\overline{\beta} = \widetilde{\beta}$ we may get

$$\lim_{n\to\infty}\boldsymbol{\beta}(\boldsymbol{\alpha}_n) = \boldsymbol{\widetilde{\beta}} = \boldsymbol{\beta}(\boldsymbol{\alpha})$$

Thus, we established that $\beta(\alpha_n) \rightarrow \beta(\alpha)$. The mapping $\alpha^{(1)} \rightarrow \beta^{(1)}$ is continuous for all sequences $\alpha_n \rightarrow \alpha$. \Box

Proof of Lemma 3. Given λ_{β} and λ_{α} , from the definition $(\hat{\boldsymbol{\beta}}^{(v+1)}, \hat{\boldsymbol{\alpha}}^{(v+1)})$ in the iterative Lasso expectile algorithm, it is easy to obtain

$$L(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v)}) \geq L(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v+1)}) \geq L(\hat{\boldsymbol{\beta}}^{(v+1)}, \hat{\boldsymbol{\alpha}}^{(v+1)})$$

Since *L* is a strictly convex function, unless $(\hat{\boldsymbol{\beta}}^{(v+1)}, \hat{\boldsymbol{\alpha}}^{(v+1)}) = (\overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\alpha}})$, at least one greater-than symbol must be included in the two formulas above. As a result, the sequence $L(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v)})$ is strictly decreasing. And since $L(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v)}) > 0$, its limit is real, indicated by \overline{L} .

And consider

$$\boldsymbol{A} = \left\{ (\boldsymbol{\beta}, \boldsymbol{\alpha}) \middle| L(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq L(\hat{\boldsymbol{\beta}}^{(0)}, \hat{\boldsymbol{\alpha}}^{(0)}) \right\}$$

is a bounded closed region.

Clearly, we can have $(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v)}) \in \boldsymbol{A}$. Denote $\lim_{h \to \infty} (\hat{\boldsymbol{\beta}}^{(v_h)}, \hat{\boldsymbol{\alpha}}^{(v_h)}) = (\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}})$ by randomly

selecting a convergent subsequence $(\hat{\beta}^{(v_h)}, \hat{\alpha}^{(v_h)})$. Because the continuity of *L*, we can obtain

$$\lim_{h\to\infty}(\hat{\boldsymbol{\beta}}^{(v_h)},\hat{\boldsymbol{\alpha}}^{(v_h)})=L(\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\alpha}})$$

Assuming that $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \neq (\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\alpha}})$, that is, $L(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) > L(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\alpha}})$, so we can further obtain $(\tilde{\boldsymbol{\beta}}^{(1)}, \tilde{\boldsymbol{\alpha}}^{(1)})$ to produce

$$r = L(\widetilde{\boldsymbol{eta}}, \widetilde{\boldsymbol{\alpha}}) - L(\widetilde{\boldsymbol{eta}}^{(1)}, \widetilde{\boldsymbol{\alpha}}^{(1)}) > 0$$

Define $(\beta^{(1)}, \alpha^{(1)}) = h(\beta, \alpha)$. It can be seen from Lemma 2 that *h* is continuous, and because $L(h(\beta, \alpha))$ and *L* are continuous, $\approx > 0$ is present, resulting in

$$\left|L(h(\boldsymbol{\beta},\boldsymbol{\alpha})) - L(h(\boldsymbol{\widetilde{\beta}},\boldsymbol{\widetilde{\alpha}}))\right| = \left|L(\boldsymbol{\beta}^{(1)},\boldsymbol{\alpha}^{(1)}) - L(\boldsymbol{\widetilde{\beta}}^{(1)},\boldsymbol{\widetilde{\alpha}}^{(1)})\right| < \frac{r}{2}$$

to make it true for every $(\boldsymbol{\beta}, \boldsymbol{\alpha}) \in U((\boldsymbol{\widetilde{\beta}}, \boldsymbol{\widetilde{\alpha}}), \boldsymbol{\asymp})_{\sim}$

It can be seen from $\lim_{h\to\infty} (\hat{\boldsymbol{\beta}}^{(v_h)}, \hat{\boldsymbol{\alpha}}^{(v_h)}) = (\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}})$ that when *h* is large enough, we have

$$L(\hat{\boldsymbol{\beta}}^{(v_h+1)}, \hat{\boldsymbol{\alpha}}^{(v_h+1)}) \leq L(\widetilde{\boldsymbol{\beta}}^{(1)}, \widetilde{\boldsymbol{\alpha}}^{(1)}) + \frac{r}{2}$$

Thus, to sum up $L(\hat{\boldsymbol{\beta}}^{(v_h+1)}, \hat{\boldsymbol{\alpha}}^{(v_h+1)}) \leq L(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}) - r + \frac{r}{2} = L(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}) - \frac{r}{2}$. This is obviously in contradiction with

$$\lim_{h\to\infty} L(\hat{\boldsymbol{\beta}}^{(v_h)}, \hat{\boldsymbol{\alpha}}^{(v_h)}) = L(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}})$$

So $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) = (\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\alpha}})$. We can obtain $(\hat{\boldsymbol{\beta}}^{(v)}, \hat{\boldsymbol{\alpha}}^{(v)}) \to (\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\alpha}})$ due to the limitation of any subsequent being the same.

The proof is ended. \Box

Proof of Theorem 1. First, similarly [27], we decompose the objective function and equation to prove Theorem 1, we define the objective function as following

$$R_{mn}(\boldsymbol{u}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_{\tau} \left\{ y_{ij} - \kappa_{ij}(\tau) - \frac{\boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{u}^{(1)}}{\sqrt{mn}} - \frac{\boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{u}_{i}^{(2)}}{\sqrt{m}} \right\} - \rho_{\tau} \left\{ y_{ij} - \kappa_{ij}(\tau) \right\}$$
(A1)

Our goal is to approximate R_{mn} by a quadratic function with a unique minimizing value, and use results to show that the asymptotic distribution of that minimizing quadratic function. This quadratic approximation is mainly composed by the Taylor expansion of the expected value and by a linear approximation function.

Let $\tilde{\mathbf{x}}_{ij} = (\mathbf{x}_{ij}^{\mathrm{T}}, \mathbf{z}_{ij}^{\mathrm{T}})^{\mathrm{T}}$, $\tilde{\mathbf{u}} = (\mathbf{u}^{(1)\mathrm{T}} / \sqrt{mn}, \mathbf{u}^{(2)\mathrm{T}} / \sqrt{m})^{\mathrm{T}}$, $r_{ij\tau} = y_{ij} - \kappa_{ij}(\tau)$, The function $\mathbb{E}(\rho_{\tau}(r_{ij\tau} - \tilde{\mathbf{x}}_{ij}^{\mathrm{T}}\tilde{\mathbf{u}}) - \rho_{\tau}(r_{ij\tau}))$ is convex, continuously differentiable twice, reaches the minimum value when $\tilde{\mathbf{u}} = 0$, and around $\tilde{\mathbf{u}} = 0$ it is represented as

$$\mathbb{E}\Big[\rho_{\tau}\Big(r_{ij\tau} - \widetilde{\mathbf{x}}_{ij}^{\mathrm{T}}\widetilde{\mathbf{u}}\Big) - \rho_{\tau}(r_{ij\tau})\Big] = \widetilde{\mathbf{u}}^{\mathrm{T}}\widetilde{\mathbf{x}}_{ij}\mathbb{E}\big[\varphi_{\tau}(r_{ij\tau})\big]\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}}\widetilde{\mathbf{u}} -2\widetilde{\mathbf{u}}^{\mathrm{T}}\widetilde{\mathbf{x}}_{ij}\mathbb{E}\big[\varphi_{\tau}(r_{ij\tau})\cdot r_{ij\tau}\big] + o\|\|\widetilde{\mathbf{u}}\|^{2}$$
(A2)

where $\varphi_{\tau}(u) = \tau - I(u < 0)$. According to

$$\underset{u \in \mathbb{R}^{n+p}}{\arg\min} \mathbb{E} \left[\rho_{\tau} \left(r_{ij\tau} - \widetilde{\mathbf{x}}_{ij}^{\mathrm{T}} \widetilde{\mathbf{u}} \right) - \rho_{\tau} \left(r_{ij\tau} \right) \right] = 0$$
(A3)

We give the first-order condition

$$\mathbb{E}\left[\varphi_{\tau}(r_{ij\tau})\cdot r_{ij\tau}\right] = 0 \tag{A4}$$

Equation (13) can be simplified as follows

$$\mathbb{E}\left[\rho_{\tau}\left(r_{ij\tau}-\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}}\widetilde{\mathbf{u}}\right)-\rho_{\tau}\left(r_{ij\tau}\right)\right]=\widetilde{\mathbf{u}}^{\mathrm{T}}\widetilde{\mathbf{x}}_{ij}\mathbb{E}\left[\varphi_{\tau}\left(r_{ij\tau}\right)\right]\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}}\widetilde{\mathbf{u}}+o\parallel\widetilde{\mathbf{u}}\parallel^{2}$$
(A5)

Taylor expansion of Equation (12) around $\tilde{u} = 0$ can be regarded as a linear approximation function. Define

$$D_{ij}(r_{ij\tau}) = -2\varphi_{\tau}(r_{ij\tau}) \cdot r_{ij\tau}$$
(A6)

According to Equation (15), and $\mathbb{E}(D_{ij}(r_{ij\tau})) = 0$. Give a definition

$$q_{ij}(\widetilde{u}) = \rho_{\tau}(r_{ij\tau} - \widetilde{x}_{ij}^{\mathrm{T}}\widetilde{u}) - \rho_{\tau}(r_{ij\tau}) - \widetilde{u}^{\mathrm{T}}\widetilde{x}_{ij}D_{ij}(r_{ij\tau})$$

And

$$R_{mn}(\widetilde{\boldsymbol{u}}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\mathbb{E} \left[\rho_{\tau}(r_{ij\tau} - \widetilde{\boldsymbol{x}}_{ij}^{T} \widetilde{\boldsymbol{u}}) - \rho_{\tau}(r_{ij\tau}) \right] \right) \\ + \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{\boldsymbol{u}}^{T} \widetilde{\boldsymbol{x}}_{ij} D_{ij}(r_{ij\tau}) + \sum_{i=1}^{n} \sum_{j=1}^{m} \left(q_{ij}(\widetilde{\boldsymbol{u}}) - \mathbb{E} \left[q_{ij}(\widetilde{\boldsymbol{u}}) \right] \right)$$

According to Koenker [5], the objective function $R_{mn}(\tilde{u})$ can be reduced to

$$R_{mn}(\widetilde{\boldsymbol{u}}) = \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\widetilde{\boldsymbol{x}}_{ij} \mathbb{E} \left[\varphi_{\tau}(r_{ij\tau}) \right] \widetilde{\boldsymbol{x}}_{ij}^{\mathrm{T}} \right) \widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{\boldsymbol{x}}_{ij} D_{ij}(r_{ij\tau}) + O \|\widetilde{\boldsymbol{u}}\|^{2} + o \|\widetilde{\boldsymbol{u}}\|^{2}$$

$$= \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\widetilde{\boldsymbol{x}}_{ij} \mathbb{E} \left[\varphi_{\tau}(r_{ij\tau}) \right] \widetilde{\boldsymbol{x}}_{ij}^{\mathrm{T}} \right) \widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{\boldsymbol{x}}_{ij} D_{ij}(r_{ij\tau}) + o_{p}(1)$$

$$\simeq \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\widetilde{\boldsymbol{x}}_{ij} \mathbb{E} \left[\varphi_{\tau}\left(r_{ij\tau} \right) \right] \widetilde{\boldsymbol{x}}_{ij}^{\mathrm{T}} \right) \widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{u}}^{\mathrm{T}} \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{\boldsymbol{x}}_{ij} D_{ij}(r_{ij\tau})$$
(A7)

Let
$$\widetilde{\mathbf{x}}_{ij} = \left(\mathbf{z}_{ij}^{\mathrm{T}}, \mathbf{x}_{ij}^{\mathrm{T}}\right)^{\mathrm{T}}, \widetilde{\mathbf{u}} = \left(\mathbf{u}_{0}^{\mathrm{T}}/\sqrt{m}, \mathbf{u}_{1}^{\mathrm{T}}/\sqrt{mn}\right)^{\mathrm{T}}$$
, Then there is

$$\mathbf{R}_{mn}(\mathbf{u}) = -2\frac{1}{\sqrt{m}}\sum_{i=1}^{n}\sum_{j=1}^{m} \left(\mathbf{z}_{ij}^{\mathrm{T}}\mathbf{u}_{0} + \mathbf{x}_{ij}^{\mathrm{T}}\mathbf{u}_{1}/\sqrt{n}\right)\varphi_{\tau}\left(\mathbf{y}_{ij} - \mathbf{\kappa}_{ij}(\tau)\right) \cdot \left(\mathbf{y}_{ij} - \mathbf{\kappa}_{ij}(\tau)\right)$$

$$+\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}\left[\varphi_{\tau}\left(\mathbf{y}_{ij} - \mathbf{\kappa}_{ij}(\tau)\right)\right]\left(\mathbf{z}_{ij}^{\mathrm{T}}\mathbf{u}_{0} + \mathbf{x}_{ij}^{\mathrm{T}}\mathbf{u}_{1}/\sqrt{n}\right)^{2}$$

$$= -2\frac{1}{\sqrt{m}}\left[\mathbf{u}_{0}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{\Pi}_{\tau}(r_{\tau})r_{\tau} + \mathbf{u}_{1}^{\mathrm{T}}/\sqrt{n}\mathbf{X}^{\mathrm{T}}\mathbf{\Pi}_{\tau}(r_{\tau})r_{\tau}\right]$$

$$+\frac{1}{m}\left[\mathbf{u}_{0}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})]\mathbf{Z}\mathbf{u}_{0} + 2\mathbf{u}_{0}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbb{E}[\mathbf{\Pi}_{\tau}(r_{\tau})]\mathbf{X}\mathbf{u}_{1}/\sqrt{n} + \mathbf{u}_{1}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{E}[\mathbf{\Pi}_{\tau}(r_{\tau})]\mathbf{X}\mathbf{u}_{1}/n\right]$$

$$= \Gamma_{mn}^{(1)}(\mathbf{u}) + \Gamma_{mn}^{(2)}(\mathbf{u})$$
(A8)

Therefore, we decompose $\Gamma_{mn}(u)$ into four parts

$$\Gamma_{mn}(u) = \Gamma_{mn}^{(1)}(u) + \Gamma_{mn}^{(2)}(u) + \Gamma_{mn}^{(3)}(u) + \Gamma_{mn}^{(4)}(u)$$
(A9)

where

$$\Gamma_{mn}^{(1)}(\boldsymbol{u}) = -2 \frac{1}{\sqrt{m}} \Big[\boldsymbol{u}_0^{\mathrm{T}} \mathbf{Z}^{\mathrm{T}} \boldsymbol{\Pi}_{\tau}(r_{\tau}) r_{\tau} + \boldsymbol{u}_1^{\mathrm{T}} / \sqrt{n} \mathbf{X}^{\mathrm{T}} \boldsymbol{\Pi}_{\tau}(r_{\tau}) r_{\tau} \Big]$$

$$\Gamma_{mn}^{(2)}(\boldsymbol{u}) = \frac{1}{m} \Big[\boldsymbol{u}_0^{\mathrm{T}} \mathbf{Z}^{\mathrm{T}} \mathbb{E}[\boldsymbol{\Pi}_{\tau}(r_{\tau})] \mathbf{Z} \boldsymbol{u}_0 + 2 \boldsymbol{u}_0^{\mathrm{T}} \mathbf{Z}^{\mathrm{T}} \mathbb{E}[\boldsymbol{\Pi}_{\tau}(\boldsymbol{\varepsilon}_{\tau})] \mathbf{X} \boldsymbol{u}_1 / \sqrt{n} + \boldsymbol{u}_1^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbb{E}[\boldsymbol{\Pi}_{\tau}(r_{\tau})] \mathbf{X} \boldsymbol{u}_1 / n \Big]$$

$$\Gamma_{mn}^{(3)}(\boldsymbol{u}) = \lambda_{\beta} \sum_{k=1}^{h} \left(\left| \beta_k + \frac{\boldsymbol{u}_k^{(1)}}{\sqrt{mn}} \right| - |\beta_k| \right)$$

$$\Gamma_{mn}^{(4)}(\boldsymbol{u}) = \lambda_{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{l} \left(\left| \boldsymbol{\alpha}_{it} + \frac{\boldsymbol{u}_{it}^{(2)}}{\sqrt{m}} \right| - |\boldsymbol{\alpha}_{it}| \right)$$

For $\Gamma_{mn}^{(1)}(\boldsymbol{u})$, conditions A2 and A3 mean Lindbergh condition, we have

$$\Gamma_{mn}^{(1)}(\boldsymbol{u}) = -2\frac{1}{\sqrt{m}} \Big[\boldsymbol{u}^{(1)\mathrm{T}} \mathbf{Z}^{\mathrm{T}} + \boldsymbol{u}^{(2)\mathrm{T}} / \sqrt{n} \mathbf{X}^{\mathrm{T}} \Big] \mathbf{\Pi}_{\tau}(r_{\tau}) r_{\tau} \stackrel{d}{\to} -2\boldsymbol{u}^{\mathrm{T}} \mathbf{W}$$
(A10)

where $W \sim N(0, P_0)$.

For $\Gamma_{mn}^{(2)}(u)$, according to hypothesis A2, we have

$$\Gamma_{nm}^{(2)}(u) = \frac{1}{m} \Big[u^{(1)T} Z^{T} \mathbb{E}[\Pi_{\tau}(r_{\tau})] Z u^{(1)} + 2u^{(1)T} Z^{T} \mathbb{E}[\Pi(r_{\tau})] X u_{1} / \sqrt{n} + u_{1}^{T} X^{T} \mathbb{E}[\Pi(r_{\tau})] X u_{1} / n \Big]$$

$$\to u^{T} P_{1} u$$
(A11)

For $\Gamma_{mn}^{(3)}(\boldsymbol{u})$,

by

$$\Gamma_{mn}^{(3)}(\boldsymbol{u}) = \lambda_{\beta} \sum_{k=1}^{h} \left(\left| \beta_{k} - \frac{u_{k}^{(1)}}{\sqrt{mn}} \right| - \left| \beta_{k} \right| \right)$$

$$= \frac{\lambda_{\beta}}{\sqrt{mn}} \sum_{k=1}^{h} \left[u_{k}^{(1)} \operatorname{sgn}(\beta_{k}) I(\beta_{k} \neq 0) + \left| u_{k}^{(1)} \right| I(\beta_{k} = 0) \right]$$

$$\to \lambda_{1} \sum_{h} \left[u_{k}^{(1)} \operatorname{sgn}(\beta_{k}) I(\beta_{k} \neq 0) + \left| u_{k}^{(1)} \right| I(\beta_{k} = 0) \right]$$
(A12)

For $\Gamma_{mn}^{(4)}(u)$, it can be divided into two parts

$$\Gamma_{mn}^{(4)}(u) = \Gamma_{mn}^{(41)}(u) + \Gamma_{mn}^{(42)}(u)$$

$$|u - v| - |u| = -v \operatorname{sgn}(u) + 2\int_0^v (I(u \le s) - I(u \le 0)) ds.$$
(A13)

where $\Gamma_{mn}^{(41)}(\boldsymbol{u}) = \lambda_{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{l} \frac{u_{it}^{(2)}}{\sqrt{m}} \operatorname{sgn}(\alpha_{it}) = \frac{\lambda_{\alpha}}{\sqrt{m}} \sum_{i=1}^{n} \sum_{t=1}^{l} u_{it}^{(2)} \sum_{i=1}^{n} \sum_{t=1}^{l} u_{it}^{(2)} \operatorname{sgn}(\alpha_{it}) \to \lambda_2 S^{\mathrm{T}} \boldsymbol{u}^{(2)}$ and $S \sim N(0, diag(4G_{it}(0)(1 - G_{it}(0))));$

$$\Gamma_{mn}^{(42)}(u) = 2\frac{\lambda_{\alpha}}{\sqrt{m}}\sum_{i=1}^{n}\sum_{t=1}^{l}\int_{0}^{-u_{it}^{(2)}}\left(I\left(\alpha_{it} \leq \frac{s}{\sqrt{m}}\right) - I(\alpha_{it} \leq 0)\right)ds$$

In hypothesis A2, $\exists a > 0$, $n^a/m \to 0$ is satisfied, we have

$$E\left(\Gamma_{mn}^{(42)}(\boldsymbol{u})\right) = 2\frac{\lambda_{\alpha}}{\sqrt{m}}\sum_{i=1}^{n}\sum_{t=1}^{l}\int_{0}^{-u_{it}^{(2)}}E\left[I\left(\alpha_{it} \leq \frac{s}{\sqrt{m}}\right) - I(\alpha_{it} \leq 0)\right]ds$$
$$= 2\frac{\lambda_{\alpha}}{\sqrt{m}}\sum_{i=1}^{n}\sum_{t=1}^{l}\int_{0}^{-u_{it}^{(2)}}\left[G_{it}\left(\frac{s}{\sqrt{m}}\right) - G_{it}(0)\right]ds$$
$$= \frac{\lambda_{\alpha}}{m}\sum_{i=1}^{n}\sum_{t=1}^{l}g_{it}(0)\left(u_{it}^{(2)}\right)^{2} + o(1)$$
$$\rightarrow 0$$

And $\operatorname{Var}\left(\Gamma_{mn}^{(42)}(\boldsymbol{u})\right) \to 0$.

According to the results of Equations (A1)–(A13). Although the function $\Gamma_{mn}(u)$ is convex and the point at which it reaches the minimum value is unique, we obtain

$$\hat{\boldsymbol{u}} = \operatorname{argmin}\Gamma_{mn}(\boldsymbol{u}) \xrightarrow{d} \operatorname{argmin}\Gamma_0(\boldsymbol{u})$$
(A14)

The proof is ended. \Box

Appendix B

The solution paths of β and α of DLQR and DLER under the two criteria.



Figure A1. The solution paths of β and α of DLQR under criterion SIC.



Figure A2. The solution paths of β and α of DLQR under criterion GACV.



Estimation of the alpha_0



Figure A3. The solution paths of β and α of DLER under criterion SIC.



Estimation of the alpha_0



Figure A4. The solution paths of β and α of DLER under criterion GACV.

Appendix C

Table A1. The results 1 of simulation 1 under three signal-to-noise ratio.

Parameters	Method	MSE	SD _{int}	\mathbf{SD}_1	SD_2	SD_3	SD_4	SD_5	SD_6	SD_7	SD_8
$\tau = 0.25$											
	DLER-SIC	0.306	0.300	0.313	0.288	0.029	0.033	0.054	0.022	0.027	0.012
_ 1	DLER-GACV	0.486	0.303	0.411	0.383	0.027	0.035	0.056	0.024	0.035	0.025
$\sigma = 1$	DLQR-SIC	0.280	0.261	0.297	0.250	0.060	0.046	0.091	0.015	0.024	0.020
	DLQR-GACV	0.289	0.266	0.299	0.254	0.063	0.049	0.088	0.037	0.039	0.031
	DLER-SIC	0.388	0.261	0.307	0.292	0.063	0.070	0.113	0.074	0.079	0.080
2	DLER-GACV	0.490	0.266	0.371	0.342	0.066	0.069	0.113	0.085	0.084	0.078
$\sigma = 2$	DLQR-SIC	0.527	0.234	0.332	0.335	0.108	0.088	0.167	0.076	0.069	0.086
	DLQR-GACV	0.535	0.236	0.333	0.324	0.110	0.098	0.169	0.083	0.086	0.084
- 2	DLER-SIC	0.585	0.337	0.324	0.321	0.117	0.129	0.178	0.138	0.107	0.113
$\sigma = 3$	DLER-GACV	0.649	0.342	0.330	0.343	0.139	0.133	0.186	0.139	0.128	0.110
	DLQR-SIC	0.868	0.341	0.337	0.375	0.162	0.162	0.244	0.134	0.165	0.117
	DLQR-GACV	0.894	0.353	0.325	0.398	0.163	0.163	0.244	0.153	0.152	0.119
$\tau = 0.5$											
	DLER-SIC	0.292	0.283	0.305	0.265	0.016	0.015	0.048	0.013	0.013	0
	DLER-GACV	0.509	0.283	0.438	0.342	0.015	0.019	0.049	0.014	0	0
$\sigma = 1$	DLQR-SIC	0.207	0.242	0.262	0.224	0.026	0.022	0.083	0.015	0.015	0.017
	DLQR-GACV	0.219	0.240	0.256	0.231	0.032	0.038	0.093	0.029	0.026	0.032
	DLER-SIC	0.403	0.278	0.284	0.335	0.072	0.075	0.128	0.067	0.081	0.071
2	DLER-GACV	0.472	0.276	0.320	0.359	0.076	0.074	0.133	0.065	0.069	0.063
$\sigma = 2$	DLQR-SIC	0.424	0.273	0.262	0.325	0.088	0.095	0.160	0.079	0.058	0.073
	DLQR-GACV	0.464	0.272	0.269	0.324	0.095	0.097	0.167	0.086	0.075	0.091
- 2	DLER-SIC	0.558	0.333	0.313	0.326	0.109	0.095	0.171	0.106	0.121	0.111
$\sigma = 3$	DLER-GACV	0.634	0.334	0.323	0.338	0.129	0.125	0.179	0.134	0.126	0.115
	DLQR-SIC	0.673	0.317	0.300	0.346	0.119	0.135	0.216	0.122	0.106	0.114
	DLQR-GACV	0.720	0.320	0.311	0.361	0.118	0.129	0.218	0.133	0.121	0.121
$\tau = 0.75$											
	DLER-SIC	0.374	0.272	0.322	0.316	0.021	0.027	0.056	0.012	0.017	0
4	DLER-GACV	0.456	0.275	0.389	0.369	0.025	0.020	0.058	0	0.021	0.007
$\sigma = 1$	DLQR-SIC	0.307	0.275	0.320	0.255	0.054	0.027	0.105	0.024	0.023	0.028
	DLQR-GACV	0.321	0.275	0.318	0.260	0.070	0.039	0.097	0.031	0.039	0.041
	DLER-SIC	0.454	0.263	0.321	0.302	0.071	0.087	0.129	0.098	0.083	0.056
2	DLER-GACV	0.538	0.262	0.400	0.359	0.072	0.086	0.124	0.106	0.070	0.064
$\sigma = 2$	DLQR-SIC	0.547	0.251	0.319	0.309	0.097	0.117	0.187	0.093	0.060	0.070
	DLQR-GACV	0.570	0.255	0.317	0.317	0.103	0.117	0.198	0.099	0.081	0.078
	DLER-SIC	0.577	0.312	0.330	0.318	0.116	0.145	0.183	0.111	0.119	0.125
- 2	DLER-GACV	0.617	0.309	0.347	0.336	0.130	0.146	0.187	0.116	0.123	0.115
$\sigma = 3$	DLQR-SIC	0.701	0.283	0.336	0.322	0.119	0.137	0.246	0.131	0.129	0.129
	DLQR-GACV	0.770	0.304	0.329	0.348	0.121	0.153	0.247	0.154	0.150	0.157

Parameters	Method	Corr (SD)	Incorr (SD)	Int	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$\tau = 0.25$												
	DLER-SIC	3 (0)	1.04 (0.665)	14	100	100	95	95	100	97	96	99
_ 1	DLER-GACV	2.99 (0.1)	1.11 (0.751)	14	100	99	96	94	100	95	94	96
$\sigma = 1$	DLQR-SIC	3 (0)	1.29 (0.518)	0	100	100	87	91	100	99	96	98
	DLQR-GACV	3 (0)	1.5 (0.628)	0	100	100	84	91	100	92	90	93
	DLER-SIC	3 (0)	1.6 (0.865)	1	100	100	90	88	100	88	88	85
- 0	DLER-GACV	2.99 (0.1)	1.81 (0.982)	1	100	99	87	84	100	82	82	83
$\sigma = 2$	DLQR-SIC	3 (0)	1.99 (0.959)	0	100	100	73	79	100	82	87	80
	DLQR-GACV	3 (0)	2.04 (1.063)	0	100	100	75	76	100	83	81	81
~ 2	DLER-SIC	3 (0)	1.87 (1.116)	0	100	100	85	83	100	83	79	83
$v \equiv s$	DLER-GACV	3 (0)	2.24 (1.311)	0	100	100	73	76	100	78	72	77
	DLQR-SIC	3 (0)	2.56 (1.351)	0	100	100	67	70	100	69	65	73
	DLQR-GACV	3 (0)	2.41 (1.280)	0	100	100	64	70	100	72	74	79
$\tau = 0.5$												
	DLER-SIC	3 (0)	0.75 (0.520)	29	100	100	99	99	100	99	99	100
_ 1	DLER-GACV	2.98 (0.2)	0.74 (0.525)	30	99	99	99	98	100	99	100	100
$\sigma = 1$	DLQR-SIC	3 (0)	0.78 (0.645)	35	100	100	96	96	100	98	99	98
	DLQR-GACV	3 (0)	0.95 (0.770)	36	100	100	95	91	100	94	96	93
	DLER-SIC	2.99 (0.1)	1.37 (0.960)	29	100	99	89	87	100	86	85	87
- 0	DLER-GACV	2.99 (0.1)	1.38 (1.080)	27	100	99	85	86	100	89	86	89
$v \equiv z$	DLQR-SIC	2.99(0.1)	1.66 (1.075)	27	100	99	78	78	100	83	88	80
	DLQR-GACV	2.99 (0.1)	1.76 (1.264)	31	100	99	76	78	100	79	81	79
$\sigma = 2$	DLER-SIC	3 (0)	1.7 (1.124)	25	100	100	82	79	100	80	81	83
v = s	DLER-GACV	3 (0)	1.98 (1.303)	28	100	100	74	71	100	73	78	78
	DLQR-SIC	3 (0)	2.38 (1.153)	22	100	100	66	69	100	62	72	71
	DLQR-GACV	3 (0)	2.22 (1.211)	20	100	100	70	69	100	74	70	75
$\tau = 0.75$												
	DLER-SIC	2.99 (0.1)	0.94 (0.528)	15	100	99	98	96	100	99	98	100
_ 1	DLER-GACV	2.99 (0.1)	0.96 (0.470)	14	100	99	96	98	100	100	97	99
$\sigma = 1$	DLQR-SIC	3 (0)	1.29 (0.478)	0	100	100	87	96	100	96	98	94
	DLQR-GACV	3 (0)	1.47 (0.717)	0	100	100	80	94	100	94	93	92
	DLER-SIC	3 (0)	1.72 (0.877)	0	100	100	88	84	100	82	86	88
~)	DLER-GACV	2.99 (0.1)	1.84 (1.042)	1	100	99	83	85	100	75	88	84
$\sigma = 2$	DLQR-SIC	3 (0)	2.04 (0.974)	0	100	100	78	72	100	71	92	83
	DLQR-GACV	3 (0)	2.12 (1.113)	0	100	100	74	75	100	75	85	79
	DLER-SIC	3 (0)	1.93 (1.066)	0	100	100	87	80	100	83	82	75
~ 2	DLER-GACV	3 (0)	2.16 (1.245)	0	100	100	78	77	100	78	76	75
v = 3	DLQR-SIC	3 (0)	2.44 (1.192)	0	100	100	72	69	100	70	72	73
	DLQR-GACV	3 (0)	2.62 (1.237)	0	100	100	67	70	100	69	68	64

 Table A2. The results 2 of simulation 1 under three signal-to-noise ratio.

τ= 0.5	Method	MSE	SD _{int}	\mathbf{SD}_1	SD_2	SD_3	SD_4	SD_5	SD_6	SD_7	SD_8
P_1											
	DLER-SIC	0.336	0.268	0.314	0.281	0.016	0.012	0.054	0.022	0.010	0.026
	DLER-GACV	0.490	0.266	0.409	0.321	0.019	0.017	0.057	0.021	0.016	0.019
	DLQR-SIC	0.286	0.229	0.294	0.279	0.051	0.037	0.079	0.030	0.021	0.018
	DLQR-GACV	0.294	0.233	0.298	0.276	0.056	0.043	0.085	0.038	0.034	0.031
P_2											
	DLER-SIC	1.318	0.539	0.559	0.568	0.013	0	0.049	0.021	0.014	0.012
	DLER-GACV	2.231	0.538	0.789	0.574	0.020	0	0.051	0.020	0.013	0.011
	DLQR-SIC	0.898	0.520	0.534	0.508	0.045	0.478	0.099	0.022	0.026	0
	DLQR-GACV	0.916	0.517	0.528	0.516	0.052	0.049	0.107	0.041	0.037	0.018
P ₃											
	DLER-SIC	2.577	0.876	0.906	0.726	0.022	0.024	0.049	0	0.017	0.022
	DLER-GACV	3.946	0.879	1.065	0.733	0.015	0.018	0.052	0.014	0.013	0.012
	DLQR-SIC	1.869	0.732	0.793	0.746	0.040	0.092	0.108	0.027	0.011	0.010
	DLQR-GACV	1.878	0.727	0.803	0.760	0.050	0.099	0.076	0.027	0.019	0.020

Table A3. The results 1 of simulation 2 under different influence of random effects.

Table A4. The results 2 of simulation 2 with different influence of random effects.

$\tau = 0.5$	Method	Corr (SD)	Incorr (SD)	Int	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
P_1												
	DLER-SIC	3 (0)	0.8 (0.651)	29	100	100	98	99	100	98	99	97
	DLER-GACV	3 (0)	0.81 (0.598)	29	100	100	98	98	100	98	98	98
	DLQR-SIC	3 (0)	1.03 (0.643)	30	100	100	85	93	100	94	97	98
	DLQR-GACV	3 (0)	1.17 (0.877)	28	100	100	85	90	100	91	94	95
<i>P</i> ₂												
	DLER-SIC	2.91 (0.288)	0.98 (0.376)	9	100	91	99	100	100	97	98	99
	DLER-GACV	2.78 (0.462)	0.99 (0.362)	8	98	80	98	100	100	98	98	99
	DLQR-SIC	3 (0)	1.12 (0.591)	15	15	100	90	90	100	98	96	100
	DLQR-GACV	3 (0)	1.24 (0.754)	17	17	100	87	90	100	91	93	98
P ₃												
	DLER-SIC	2.75 (0.435)	1.04 (0.470)	6	99	76	98	98	100	100	98	96
	DLER-GACV	2.46 (0.673)	1.01 (0.460)	7	90	56	99	98	100	99	98	98
	DLQR-SIC	2.97 (0.171)	1.11 (0.601)	15	100	100	93	87	100	96	99	99
	DLQR-GACV	2.97 (0.171)	1.24 (0.605)	4	100	100	87	84	100	96	98	97

Table A5. The results 1 of simulation 3 with dense model under different error distributions.

Model	Method	MSE	SD_{int}	SD_1	SD_2	SD_3	SD_4	SD_5	SD_6	SD_7	SD_8
Dense											
	DLER-SIC	0.908	0.522	0.544	0.527	0.058	0.061	0.066	0.068	0.058	0.053
$\mathbf{N}(0, 1)$	DLER-GACV	0.922	0.522	0.558	0.550	0.058	0.062	0.066	0.068	0.059	0.053
N(0,1)	DLQR-SIC	0.853	0.468	0.485	0.527	0.079	0.079	0.081	0.093	0.076	0.070
	DLQR-GACV	0.882	0.463	0.476	0.536	0.088	0.086	0.087	0.102	0.084	0.073
	DLER-SIC	0.965	0.556	0.570	0.534	0.140	0.097	0.104	0.113	0.108	0.113
4(2)	DLER-GACV	1.114	0.551	0.560	0.590	0.187	0.114	0.125	0.126	0.131	0.163
l(3)	DLQR-SIC	0.871	0.518	0.525	0.467	0.113	0.099	0.101	0.118	0.108	0.094
	DLQR-GACV	0.872	0.520	0.520	0.484	0.115	0.095	0.088	0.113	0.103	0.095
	DLER-SIC	4.397	0.391	0.404	0.678	0.105	0.272	0.391	0.370	0.753	0.377
C_{1}	DLER-GACV	4.828	0.535	0.405	0.795	0.306	0.310	0.397	0.370	0.725	0.424
Cuucny(0,1)	DLQR-SIC	2.450	0.249	0.350	0.605	0.039	0.089	0.031	0.072	0.092	0.127
	DLQR-GACV	2.157	0.237	0.361	0.555	0.008	0.0536	0.041	0.020	0.094	0.129

Model	Method	Corr (SD)	Incorr (SD)	Int	X_1	<i>X</i> ₂	X_3	X_4	X_5	X_6	X_7	X_8
Dense												
	DLER-SIC	7.99 (0.33)	0.74 (0.44)	26	94	96	100	100	100	100	100	100
N(0,1)	DLER-GACV	7.7 (0.63)	0.74 (0.44)	26	84	86	100	100	100	100	100	100
N(0,1)	DLQR-SIC	7.9 (0.33)	0.78 (0.42)	22	94	96	100	100	100	100	100	100
	DLQR-GACV	7.91 (0.32)	0.81 (0.39)	19	96	95	100	100	100	100	100	100
	DLER-SIC	7.72 (0.62)	0.84 (0.37)	16	83	90	99	100	100	100	100	100
t(2)	DLER-GACV	7.37 (0.92)	0.83 (0.38)	17	66	73	99	100	100	100	100	99
1(3)	DLQR-SIC	7.88 (0.33)	0.84 (0.37)	16	93	95	100	100	100	100	100	100
	DLQR-GACV	7.91 (0.29)	0.84 (0.37)	16	95	96	100	100	100	100	100	100
	DLER-SIC	7.6 (0.80)	0.8 (0.40)	20	100	100	100	100	80	80	100	100
Cauchu(0,1)	DLER-GACV	7.68 (0.47)	1 (0)	0	100	100	100	100	88	80	100	100
Cuuchy(0,1)	DLQR-SIC	8 (0)	0.8 (0.40)	20	100	100	100	100	100	100	100	100
	DLQR-GACV	8 (0)	0.8 (0.40)	20	100	100	100	100	100	100	100	100

Table A6. The results 2 of simulation 3 with dense model under different error distributions.

Table A7. The results 1 of simulation 3 with the sparse model under different error distributions.

Model	Method	MSE	SD _{int}	SD_1	SD_2	SD ₃	SD_4	SD_5	SD_6	SD_7	SD_8
Sparse											
-	DLER-SIC	1.345	0.608	0.622	0.550	0.011	0	0.051	0.015	0.015	0.023
N(0, 1)	DLER-GACV	2.084	0.608	0.849	0.606	0.011	0	0.052	0.019	0.014	0.018
N(0,1)	DLQR-SIC	0.978	0.508	0.527	0.544	0.030	0.033	0.123	0.030	0.020	0.013
	DLQR-GACV	0.947	0.516	0.527	0.529	0.036	0.037	0.106	0.037	0.028	0.027
	DLER-SIC	1.230	0.676	0.533	0.605	0.055	0.039	0.104	0.075	0.049	0.064
4(2)	DLER-GACV	2.737	0.670	1.083	0.742	0.070	0.053	0.204	0.082	0.057	0.075
l(3)	DLQR-SIC	1.008	0.522	0.498	0.530	0.065	0.049	0.102	0.041	0.047	0.048
	DLQR-GACV	1.000	0.529	0.492	0.531	0.066	0.055	0.106	0.046	0.047	0.049
	DLER-SIC	5.756	0.620	0.732	0.404	0.121	0	0.399	0.587	0.238	0.385
Cauchu(0,1)	DLER-GACV	6.791	0.708	0.753	0.307	0.384	0.073	0.542	0.682	0.521	0.460
Cuuchy(0,1)	DLQR-SIC	2.099	0.593	0.411	0.310	0.105	0.193	0.152	0.119	0.046	0.062
	DLQR-GACV	1.758	0.638	0.385	0.342	0	0.173	0.113	0.095	0.049	0.034

Table A8. The results 2 of simulation 3 with the sparse model under different error distributions.

Model	Method	Corr (SD)	Incorr (SD)	Int	X_1	<i>X</i> ₂	<i>X</i> ₃	X_4	X_5	X_6	X_7	X_8
Sparse												
-	DLER-SIC	2.91 (0.288)	0.98 (0.40)	8	100	91	99	100	100	99	98	98
N(0, 1)	DLER-GACV	2.79 (0.478)	0.98 (0.40)	8	97	82	99	100	100	97	99	99
N(0,1)	DLQR-SIC	3 (0)	1.09 (0.62)	14	100	100	94	92	100	95	97	99
	DLQR-GACV	3 (0)	1.17 (0.63)	14	100	100	94	92	100	92	95	96
	DLER-SIC	2.94 (0.24)	1.38 (0.84)	8	100	94	91	96	100	87	93	87
4(2)	DLER-GACV	2.59 (0.71)	1.55 (1.06)	12	87	72	84	91	100	83	90	85
l(3)	DLQR-SIC	2.97 (0.17)	1.27 (0.79)	15	100	97	88	91	100	92	94	93
	DLQR-GACV	2.97 (0.17)	1.38 (0.85)	16	100	97	86	86	100	90	93	91
	DLER-SIC	2.84 (0.37)	2.02 (0.62)	0	100	84	79	100	100	80	59	80
$C_{au} = h_{u}(0, 1)$	DLER-GACV	3 (0)	4.62 (0.83)	7	100	100	2	78	100	18	15	18
Cauchy(0,1)	DLQR-SIC	3 (0)	2.66 (1.11)	0	100	100	79	36	100	71	71	77
	DLQR-GACV	3 (0)	2.09 (0.93)	0	100	100	100	43	100	71	84	93

Model	Method	MSE	SD _{int}	SD_1	SD_2	SD ₃	SD_4	SD_5	SD_6	SD ₇	SD_8
High Sparse											
Ŭ Î	DLER-SIC	1.017	0.611	0.593	0.267	0.020	0.021	0.025	0.039	0.035	0.017
N(0, 1)	DLER-GACV	3.096	0.613	0.981	0.253	0.017	0.022	0.025	0.033	0.016	0
N(0,1)	DLQR-SIC	0.773	0.523	0.493	0.475	0.038	0.029	0.030	0.020	0.018	0.035
	DLQR-GACV	0.81	0.528	0.509	0.477	0.048	0.040	0.036	0.039	0.041	0.036
	DLER-SIC	1.264	0.528	0.73	0.282	0.055	0	0.024	0.042	0.083	0.091
+(2)	DLER-GACV	2.478	0.519	1.019	0.307	0.069	0.039	0.03	0.039	0.092	0.087
l(3)	DLQR-SIC	0.792	0.474	0.576	0.394	0.082	0.072	0.051	0.061	0.036	0.039
	DLQR-GACV	0.777	0.483	0.575	0.374	0.056	0.054	0.047	0.038	0.029	0.035
	DLER-SIC	1.925	1.147	0.162	0.741	0	0.513	0	0	0	0.371
Cauchu(0,1)	DLER-GACV	14.720	1.200	0.298	1.289	0.932	1.201	0.602	0.513	0.227	0.334
Cuuchy(0,1)	DLQR-SIC	0.798	0.650	0.256	0.484	0.117	0.095	0.082	0.029	0.124	0.121
	DLQR-GACV	0.822	0.567	0.238	0.415	0.116	0.088	0.060	0.0315	0.109	0.109

 Table A9. The results 1 of simulation 3 with the high sparse model under different error distributions.

Table A10. The results 2 of simulation 3 with the high sparse model under different error distributions.

Model	Method	Corr (SD)	Incorr (SD)	Int	X_1	<i>X</i> ₂	<i>X</i> ₃	X_4	X_5	X_6	X_7	X_8
High Sparse												
	DLER-SIC	1 (0)	1.43 (0.78)	4	100	75	98	98	98	92	95	97
N(0,1)	DLER-GACV	1 (0)	1.27 (0.72)	4	100	83	98	98	98	94	98	100
N(0,1)	DLQR-SIC	1 (0)	1.59 (0.78)	29	100	35	93	97	95	99	98	95
	DLQR-GACV	1 (0)	1.71 (0.97)	31	100	39	92	92	94	95	92	94
	DLER-SIC	1 (0)	1.52 (0.92)	2	100	75	91	100	98	94	93	95
4(2)	DLER-GACV	0.99 (0.1)	1.7 (1.11)	2	99	70	87	94	96	96	94	91
l(3)	DLQR-SIC	1 (0)	2.25 (1.22)	14	100	34	78	84	89	90	94	92
	DLQR-GACV	1 (0)	1.94 (1.08)	11	100	45	85	90	90	95	96	94
	DLER-SIC	1 (0)	1.71 (0.76)	0	100	82	100	53	100	100	100	94
Cauchu(0,1)	DLER-GACV	1 (0)	6.53 (2.35)	0	100	11	12	11	29	30	30	24
Cuuchy(0,1)	DLQR-SIC	1 (0)	3.93 (1.28)	6	100	6	40	68	54	99	67	67
	DLQR-GACV	1 (0)	3.06 (1.54)	6	100	48	39	68	96	99	67	71

Table A11. The results 1 of simulation 4 with the high dimensional model.

Model	Method	MSE	SD_0	SD_1	SD ₂	SD ₅
au = 0.5						
	DLER-SIC	4.783	0.010	0.513	0.360	0.657
	DLER-GACV	5.414	0.007	0.914	0.482	0.706
	DLQR-SIC	39.235	0.406	0.585	0.606	0.443
	DLQR-GACV	6.902	0.003	0.876	0.463	0.789
$\tau = 0.75$						
	DLER-SIC	5.870	0.165	0.671	0.383	0.882
	DLER-GACV	4.483	0.010	0.759	0.467	0.719
	DLQR-SIC	40.62433	0.413	0.593	0.625	0.507
	DLQR-GACV	7.799	0.009	0.916	0.514	0.723

Model	Method	Corr (SD)	Incorr (SD)	X_0	X_1	X_2	X_5
$\tau = 0.5$							
	DLER-SIC	2.34 (0.590)	5.31 (2.237)	95.953	100	44	90
	DLER-GACV	2.28 (0.805)	3.46 (1.904)	97.682	88	50	90
	DLQR-SIC	3 (0)	94.3 (3.033)	12.794	100	100	100
	DLQR-GACV	2.8 (0.426)	68.44 (33.202)	36.972	98	91	91
$\tau = 0.75$							
	DLER-SIC	2.27 (0.737)	5.54 (7.885)	95.738	96	53	78
	DLER-GACV	2.58 (0.684)	5.32 (3.490)	95.963	96	72	90
	DLQR-SIC	2.99 (0.1)	94.46 (3.963)	12.654	100	100	99
	DLQR-GACV	2.68 (0.490)	42.06 (41.702)	61.617	98	82	88

Table A12. The results 2 of simulation 4 with the high dimensional model.

Table A13. CD4 cell data: the estimation of DLER-SIC, DLER-GACV, DLQR-SIC, DLQR-GACV.

	Int	Time X1	Smoking X ₂	Age X3	Sex Partner X ₄
0.1		1	2	5	4
$\tau = 0.1$ DI FR-SIC	_0.255	_0 271	0.080	0	0
DI FR-CACV	-0.200	-0.271	0.058	0	0
DLER GACV	-0.851	-0.314	0.030	0	0
DLOR-GACV	-0.851	-0.314	0.119	0	0
0.05	0.002	0.010			
$\tau = 0.25$	0.047	0.042	0.000	0	0
DLEK-SIC	-0.047	-0.243	0.080	0	0
DLEK-GACV	0	-0.234	0.051	0	0
DLQK-SIC	-0.422	-0.258	0.117	0	0
DLQK-GACV	-0.420	-0.259	0.123	0	0
au=0.5					
DLER-SIC	0.129	-0.233	0.090	0	0
DLER-GACV	0.112	-0.241	0.101	0	0
DLQR-SIC	0.126	-0.208	0.113	0	0
DLQR-GACV	0.126	-0.207	0.105	0	0
$\tau = 0.75$					
DLER-SIC	0.345	-0.220	0.074	0	0
DLER-GACV	0.329	-0.230	0.098	0	0
DLQR-SIC	0.557	-0.167	0.104	0	0
DLQR-GACV	0.559	-0.166	0.129	0	0
$\tau = 0.9$					
DLER-SIC	0.492	-0.217	0.078	0	0
DLER-GACV	0.492	-0.228	0.093	0	0
DLQR-SIC	0.959	-0.141	0.154	0	0
DLQR-GACV	0.949	-0.14	0.153	0	0

References

- 1. Koenker, R.; Bassett, G., Jr. Regression quantiles. Econom. J. Econom. Soc. 1978, 46, 33–50. [CrossRef]
- 2. Tibshirani, R. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 1996, 58, 267–288. [CrossRef]
- 3. Zou, H. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 2006, 101, 1418–1429. [CrossRef]
- Biswas, J.; Das, K. A Bayesian quantile regression approach to multivariate semi-continuous longitudinal data. *Comput. Stat.* 2021, 36, 241–260. [CrossRef]
- 5. Koenker, R. Quantile regression for longitudinal data. J. Multivar. Anal. 2004, 91, 74–89. [CrossRef]
- 6. Wang, H.J.; Feng, X.; Dong, C. Copula-based quantile regression for longitudinal data. Stat. Sin. 2019, 29, 245–264. [CrossRef]
- 7. Wu, Y.; Liu, Y. Variable selection in quantile regression. *Stat. Sin.* 2009, 19, 801–817.
- 8. Peng, L. Quantile regression for survival data. Annu. Rev. Stat. Its Appl. 2021, 8, 413. [CrossRef]
- 9. Li, H.; Liu, Y.; Luo, Y. Double Penalized Quantile Regression for the Linear Mixed Effects Model. J. Syst. Sci. Complex. 2020, 33, 2080–2102. [CrossRef]
- 10. Newey, W.K.; Powell, J.L. Asymmetric least squares estimation and testing. Econom. J. Econom. Soc. 1987, 55, 819–847. [CrossRef]

- 11. Almanjahie, I.M.; Bouzebda, S.; Kaid, Z.; Laksaci, A. Nonparametric estimation of expectile regression in functional dependent data. J. Nonparametr. Stat. 2022, 34, 250–281. [CrossRef]
- 12. Gu, Y.; Zou, H. High-dimensional generalizations of asymmetric least squares regression and their applications. *Ann. Stat.* **2016**, 44, 2661–2694. [CrossRef]
- 13. Farooq, M.; Steinwart, I. Learning rates for kernel-based expectile regression. Mach. Learn. 2019, 108, 203–227. [CrossRef]
- Schulze Waltrup, L.; Kauermann, G. Smooth expectiles for panel data using penalized splines. *Stat. Comput.* 2017, 27, 271–282. [CrossRef]
- 15. Sobotka, F.; Kauermann, G.; Schulze Waltrup, L.; Kneib, T. On confidence intervals for semiparametric expectile regression. *Stat. Comput.* **2013**, 23, 135–148. [CrossRef]
- 16. Zhao, J.; Zhang, Y. Variable selection in expectile regression. Commun. Stat.-Theory Methods 2018, 47, 1731–1746. [CrossRef]
- Liao, L.; Park, C.; Choi, H. Penalized expectile regression: An alternative to penalized quantile regression. *Ann. Inst. Stat. Math.* 2019, 71, 409–438. [CrossRef]
- Waldmann, E.; Sobotka, F.; Kneib, T. Bayesian regularisation in geoadditive expectile regression. *Stat. Comput.* 2017, 27, 1539–1553. [CrossRef]
- Xu, Q.F.; Ding, X.H.; Jiang, C.X.; Yu, K.M.; Shi, L. An elastic-net penalized expectile regression with applications. *J. Appl. Stat.* 2021, 48, 2205–2230. [CrossRef]
- 20. Farooq, M.; Steinwart, I. An SVM-like approach for expectile regression. Comput. Stat. Data Anal. 2017, 109, 159–181. [CrossRef]
- 21. Spiegel, E.; Sobotka, F.; Kneib, T. Model selection in semiparametric expectile regression. *Electron. J. Stat.* **2017**, *11*, 3008–3038. [CrossRef]
- 22. Daouia, A.; Girard, S.; Stupfler, G. Tail expectile process and risk assessment. Bernoulli 2020, 26, 531–556. [CrossRef]
- 23. Ziegel, J.F. Coherence and elicitability. Math. Financ. 2016, 26, 901–918. [CrossRef]
- 24. Rao, C.R.; Statistiker, M. Linear Statistical Inference and Its Applications; Wiley: New York, NY, USA, 1973.
- 25. Harville, D. Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Stat.* **1976**, *4*, 384–395. [CrossRef]
- 26. Robinson, G.K. That BLUP is a good thing: The estimation of random effects. Stat. Sci. 1991, 6, 15–31.
- Van Vleck, L.D.; Henderson, C.R. Estimates of genetic parameters of some functions of part lactation milk records. *J. Dairy Sci.* 1961, 44, 1073–1084. [CrossRef]
- Goldberger, A.S. Best linear unbiased prediction in the generalized linear regression model. J. Am. Stat. Assoc. 1962, 57, 369–375.
 [CrossRef]
- 29. Koenker, R.; Ng, P.; Portnoy, S. Quantile smoothing splines. Biometrika 1994, 81, 673-680. [CrossRef]
- 30. Yuan, M. GACV for quantile smoothing splines. Comput. Stat. Data Anal. 2006, 50, 813–829. [CrossRef]