

Accepted Manuscript

Multiple quantile regression analysis of longitudinal data:
Heteroscedasticity and efficient estimation

Hyunkeun Cho, Seonjin Kim, Mi-Ok Kim

PII: S0047-259X(17)30037-4

DOI: <http://dx.doi.org/10.1016/j.jmva.2017.01.009>

Reference: YJMVA 4213

To appear in: *Journal of Multivariate Analysis*

Received date: 24 May 2016



Please cite this article as: H. Cho, S. Kim, M.-O. Kim, Multiple quantile regression analysis of longitudinal data: Heteroscedasticity and efficient estimation, *Journal of Multivariate Analysis* (2017), <http://dx.doi.org/10.1016/j.jmva.2017.01.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Multiple quantile regression analysis of longitudinal data: heteroscedasticity and efficient estimation

Hyunkeun Cho, Seonjin Kim, and Mi-Ok Kim¹

Abstract

The objective of this paper is two-fold: to propose efficient estimation of multiple quantile regression analysis of longitudinal data and to develop a new test for the homogeneity of independent variable effects across multiple quantiles. Estimating multiple regression quantile coefficients simultaneously entails accommodating both association among the multiple quantiles and association among the repeated measurements of the response within subjects. We formulate simultaneous estimating equations using basis matrix expansion which accounts for the above-mentioned associations. The empirical likelihood method is adopted to estimate multiple regression quantile coefficients. Theoretical results show that the proposed simultaneous estimation is asymptotically more efficient than separate estimation of individual regression quantiles or ignoring the within-subject dependency. The proposed method also offers an empirical likelihood ratio test examining the homogeneity of the independent variable effects across the multiple quantiles. The Wilk's theorem holds for the test statistic, and thus the test is easy to implement. Simulation studies and real data example of a multi-center AIDS cohort study are included to illustrate the proposed estimation and testing methods, and empirically examine their properties.

Key words and phrases: Asymptotic efficiency, empirical likelihood, heteroscedasticity test, longitudinal data, multiple quantiles.

¹Hyunkeun Cho is Assistant Professor, Department of Statistics, Western Michigan University, Kalamazoo, MI 49008 (E-mail: hyunkeun.cho@wmich.edu). Seonjin Kim is Assistant Professor, Department of Statistics, Miami University, Oxford, OH 45056 (Email: kims20@miamioh.edu). Mi-Ok Kim is Professor, Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143 (Email: miok.kim@ucsf.edu).

1 Introduction

In longitudinal data, the temporal relationship between the response and explanatory variables can vary across the conditional distribution of the response when heteroscedasticity exists. For this reason, quantile regression has recently attracted attention; see Jung [6], He et al. [4], Koenker [10], Tang and Leng [18], Wang and Zhu [20], Tang et al. [19], and others. Quantile regression enables one to postulate varying effects of the independent variables across the conditional distribution. Given $\tau_1, \dots, \tau_K \in (0, 1)$, the quantiles of repeatedly measured response variables conditioning on the independent variables are given, for each $i \in \{1, \dots, n\}$, by

$$Q_{\tau_k}(y_i|x_i) = x_i\beta_{\tau_k},$$

where n is the number of subjects, $y_i = (y_{i1}, \dots, y_{im_i})^\top$ denotes the m_i response measurements observed within subject i , $x_i = (x_{i1}, \dots, x_{im_i})^\top$ is an $(m_i \times p)$ -dimensional matrix of the independent variables collected from the i th subject, and $\beta_{\tau_k} = (\beta_{\tau_k,1}, \dots, \beta_{\tau_k,p})^\top$ is a p -dimensional parameter vector at the τ_k th quantile level. While the quantile specific regression coefficients $\beta_{\tau_1}, \dots, \beta_{\tau_K}$ collectively capture the relationships with the independent variables, most of the existing work estimate the individual regression quantiles β_{τ_k} separately for each $k \in \{1, \dots, K\}$.

For each $k \in \{1, \dots, K\}$, let $\varphi_{\tau_k}(u)$ be the derivative of the so-called check function $\rho_{\tau_k}(u) = u\{\tau_k - \mathbf{1}(u < 0)\}$ and set $\varphi(u) = (\varphi_{\tau_1}(u)^\top, \dots, \varphi_{\tau_K}(u)^\top)^\top$. In this paper, we estimate $\beta_{\tau_1}, \dots, \beta_{\tau_K}$ simultaneously using generalized estimating equations [12] as follows:

$$\sum_{i=1}^n X_i^\top A_i^{-1/2} R_i(\delta)^{-1} A_i^{-1/2} \varphi(Y_i - X_i\beta) = 0, \quad (1)$$

Here $Y_i = (y_i^\top, \dots, y_i^\top)^\top$, $X_i = I_K \otimes x_i$ defined by a left Kronecker product operator \otimes and a $(K \times K)$ -dimensional identity matrix I_K , $\beta = (\beta_{\tau_1}^\top, \dots, \beta_{\tau_K}^\top)^\top$, $\varphi(Y_i - X_i\beta) = (\varphi_{\tau_1}(y_i - x_i\beta_1)^\top, \dots, \varphi_{\tau_K}(y_i - x_i\beta_K)^\top)^\top$, and A_i and $R_i(\delta)$ are an $(m_i K \times m_i K)$ -dimensional diagonal variance matrix and working correlation matrix of $\varphi(Y_i - X_i\beta)$, respectively.

The term $R_i(\delta)$ in (1) is used to approximate the true correlation matrix of $\varphi(Y_i - X_i\beta)$, denoted by Π_i ; it usually contains a low dimension of nuisance parameters δ associated with the within-subject correlation. The proposed simultaneous estimating equations accommodate not only the within-subject dependency commonly present in longitudinal data, but also a cross-correlation among the multiple quantiles, thereby providing a more efficient estimation. The efficiency gain, however, comes with an additional requirement of estimating nuisance parameters δ in $R_i(\delta)$. When a single quantile is concerned, Yi and He [23] obtained a more efficient estimator by estimating directly the true correlation matrix Π_i . Albeit possible, reliable estimation of Π_i is non-trivial, especially when m_i is large with a large number of nuisance parameters associated with the dependency structure, or a low or high quantile is of interest. This difficulty is exacerbated with multiple quantiles.

We propose to represent an inverse of $R_i(\delta)$ in (1) using the basis matrix expansion of Qu et al. [17] and avoid estimating the additional nuisance parameters associated with the correlation structure. The estimating equations (1) are expanded by basis matrices appropriately chosen for $R_i(\delta)$ and we use the empirical likelihood method [14] for the estimation of the regression coefficients β . Qin and Lawless [16] discussed the empirical likelihood method for generalized estimating equations; Yang and He [22] showed that the method similarly applies to quantile regression for independent data; Cho et al. [1] extended it to longitudinal data in the single quantile regression model. We show that the proposed simultaneous multiple quantile estimation approach is more efficient than the one either ignoring the inter-quantile correlation and estimating β_{τ_k} individually, or ignoring both the inter-quantile and within-subject correlation. As shown herein, simulation studies exhibit efficiency gain in finite samples with meaningful effects in a concrete application.

The proposed empirical likelihood approach also provides a test for the homogeneity of the independent variable relationship with the response across the multiple quantiles using the likelihood ratio statistic. Similar to its parametric counterpart, the test does not require estimating the covariance matrix of the quantile regression coefficient estimator. This is a highly desired property

as the covariance matrix of the quantile regression coefficient estimator involves the densities of the conditional distribution of the response at the quantiles of interest. Inference for this purpose is surprisingly less developed, even though quantile regression analysis and heteroscedasticity are closely associated. For situations involving independent data, Koenker and Bassett [11] and Furno [2] proposed tests. The tests require estimation of the covariance matrix of quantile regression estimators and the performance hinges on reliable estimation of the densities at the quantiles of interest. As for the inference of individual coefficients, we adopt the random perturbation approach [5] to approximate the empirical distributions of the regression quantile estimators.

There has been growing interest in properly aggregating information across multiple quantiles under the homogeneity assumption of quantile coefficients in order to yield more efficient estimators; see Koenker [9], Portnoy and Koenker [15], Zou and Yuan [26], Xiao and Koenker [21], Kai et al. [7], and Zhao and Xiao [24]. The proposed empirical likelihood test can be used to validate this assumption. When the homogeneity of a quantile coefficient cannot be rejected, the proposed empirical likelihood estimation may further improve the estimation by constraining the common coefficient to be the same across multiple quantiles.

The paper is organized as follows. Section 2 proposes efficient estimation and statistical inference in the multiple quantile regression model. Sections 3 and 4 illustrate the proposed procedure with various simulation studies and an application to an HIV data set, respectively. All proofs of theorems are provided in the Appendix.

2 Methodology

2.1 Estimation of multiple quantile regression

The working correlation structure $R_i(\delta)$ in (1) plays an important role in increasing estimation efficiency. It involves two pieces of informative associations, a within-subject correlation, denoted by $C_i(\delta)$, and cross-correlation among K quantiles, denoted by G . Accordingly, the working

correlation structure can be expressed in block matrix form as $R_i(\delta) = G \otimes C_i(\delta)$. Note that the elements (ℓ, k) and (k, ℓ) in G are known and specified as

$$\text{corr}\{\varphi_{\tau_k}(y_{ij} - x_{ij}\beta_{\tau_k}), \varphi_{\tau_\ell}(y_{ij} - x_{ij}\beta_{\tau_\ell})\} = \sqrt{\frac{\tau_\ell(1 - \tau_k)}{\tau_k(1 - \tau_\ell)}}$$

for $\ell < k$ at the true values of β_{τ_k} and β_{τ_ℓ} . Therefore, $R_i(\delta)^{-1}$ in (1) is ultimately determined by the type of the within-subject correlation structure $C_i(\delta)$.

Motivated by Qu et al. [17], the inverse of the within-subject correlation matrix is modeled by $C_i(\delta)^{-1} = \sum_{j=1}^q u_j U_{ij}$, where U_{i1}, \dots, U_{iq} are known basis matrices and u_1, \dots, u_q are unknown coefficients. This leads to

$$R_i(\delta)^{-1} = G^{-1} \otimes C_i(\delta)^{-1} = G^{-1} \otimes \sum_{j=1}^q u_j U_{ij} = \sum_{j=1}^q u_j G^{-1} \otimes U_{ij} = \sum_{j=1}^q u_j B_{ij}. \quad (2)$$

Note that U_{i1}, \dots, U_{iq} are determined by the structure of $C_i(\delta)$, whereas u_1, \dots, u_q correspond to the nuisance parameters δ in $C_i(\delta)$. For example, if $C_i(\delta)$ is assumed to be an AR(1) model, then $C_i(\delta)^{-1} = u_1 U_{i1} + u_2 U_{i2} + u_3 U_{i3}$, where U_{i1} is an identity matrix, U_{i2} is a symmetric matrix with 1 on the sub-diagonal entries and 0 elsewhere, U_{i3} is a symmetric matrix with 1 in elements $(1, 1)$ and (m_i, m_i) and 0 elsewhere, and u_1, u_2 and u_3 are unknown constants. Consequently, the inverse of $R_i(\delta)$ is expressed by three matrices as

$$\begin{aligned} R_i(\delta)^{-1} &= G^{-1} \otimes (u_1 U_{i1} + u_2 U_{i2} + u_3 U_{i3}) \\ &= u_1 G^{-1} \otimes U_{i1} + u_2 G^{-1} \otimes U_{i2} + u_3 G^{-1} \otimes U_{i3} = u_1 B_{i1} + u_2 B_{i2} + u_3 B_{i3}. \end{aligned}$$

By substituting a set of matrices $\{B_{i1}, \dots, B_{iq}\}$ in place of $R_i(\delta)^{-1}$, the equations (1) are

expanded to $\sum_{i=1}^n g_i(\beta) = 0$, where

$$g_i(\beta) = \begin{pmatrix} X_i^\top A_i^{-1/2} B_{i1} A_i^{-1/2} \varphi(Y_i - X_i \beta) \\ \vdots \\ X_i^\top A_i^{-1/2} B_{iq} A_i^{-1/2} \varphi(Y_i - X_i \beta) \end{pmatrix}. \quad (3)$$

Since the equations do not involve unknown coefficients u_1, \dots, u_q in (2), they do not require estimation of the working correlation structure. However, we cannot set each component in $\sum_{i=1}^n g_i(\beta)$ to zero simultaneously in estimating β , because not only the dimension of $g_i(\beta)$ exceeds the number of parameters, but also in view of the discreteness of $g_i(\beta)$ inherited from the loss function of the quantile regression. Following Qin and Lawless [16], we use the empirical likelihood method and construct the following likelihood:

$$L(\beta) = \sup \left\{ \prod_{i=1}^n w_i : \sum_{i=1}^n w_i g_i(\beta) = 0, \sum_{i=1}^n w_i = 1, 0 \leq w_i \leq 1 \right\},$$

where w_i denotes a point mass assigned to subject i . The maximum empirical likelihood estimator is obtained as

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} L(\beta).$$

We denote the true parameter by β_0 and assume the following conditions to study the asymptotic properties of $\hat{\beta}$.

Condition 1. Given $R_i(\delta)$, $E\{g_i(\beta_0)\} = 0$, $E\{g_i(\beta_0)g_i(\beta_0)^\top\}$ is positive definite, and $E\{\partial g_i(\beta)/\partial \beta\}$ is of full rank at $\beta = \beta_0$.

Condition 2. The conditional distribution function of Y_{ij} given x_{ij} , denoted F_{ij} , is twice continuously differentiable with bounded derivatives in the neighborhood of the quantiles of interest uniformly in x_{ij} for all i and j .

Condition 3. The random vector x_{ij} is bounded in probability for all i and j .

These are standard conditions commonly assumed for quantile regression. We denote the τ_k th

quantile of the conditional distribution of Y_{ij} given x_{ij} by $q_{ij}(\tau_k)$ and denote the conditional density at the quantile by $f_{ij}\{q_{ij}(\tau_k)\}$. With $f_i(\tau_k) = (f_{i1}\{q_{i1}(\tau_k)\}, \dots, f_{im_i}\{q_{im_i}(\tau_k)\})^\top$, we define

$$\begin{aligned}\Delta_i &= \text{diag}\{f_i(\tau_1), \dots, f_i(\tau_K)\}, \quad \Omega = E\{g_i(\beta_0)g_i(\beta_0)^\top\}, \\ \Gamma^\top &= E\left(-X_i^\top A^{-1/2} B_{i1} A^{-1/2} \Delta_i X_i, \dots, -X_i^\top A^{-1/2} B_{iq} A^{-1/2} \Delta_i X_i\right). \end{aligned} \quad (4)$$

Theorem 1. *Under regularity conditions 1–3, $\sqrt{n}(\hat{\beta} - \beta_0) \rightsquigarrow \mathcal{N}(0, \Sigma)$ in distribution as $n \rightarrow \infty$, where $\Sigma = (\Gamma^\top \Omega^{-1} \Gamma)^{-1}$.*

Theorem 1 shows that the estimator is asymptotically normal with covariance matrix Σ . We further compare the asymptotic efficiency of the proposed multiple quantile model with the ones based on the single quantile regression model with or without incorporating the within-subject correlation. Ignoring the inter-quantile correlation and both the inter-quantile and within-subject correlation corresponds to assuming working independence respectively as $G = I_K$, and $G = I_K$ and $C_i(\delta) = I_{m_i}$ in (2). We let Σ_{C_i} and Σ_I be the asymptotic covariance matrices of $\hat{\beta}$ accordingly. Here, we adopt the notations $M_1 < M_2$ and $M_1 \leq M_2$ for square matrices, M_1 and M_2 , of the same order, when $M_2 - M_1$ is positive and semi-positive definite.

Theorem 2. *If regularity conditions 1–3 hold, then $\Sigma < \Sigma_{C_i} \leq \Sigma_I$.*

Theorem 2 confirms that estimation efficiency is improved by incorporating the associations among K quantiles, as well as the within-subject dependency commonly existing in the longitudinal data. As $\Sigma < \Sigma_{C_i}$, the asymptotic covariance obtained from the multiple quantile regression model is always smaller than the one obtained from the single quantile regression model. Even though only one quantile regression is of particular interest, the proposed simultaneous multiple quantile estimation may also be desirable if simultaneous estimation is supported by a reasonably sized sample. For example, a more efficient median regression estimator can be obtained by estimating the 25th and 75th quantile regression estimators together. Moreover, the efficiency gain does not require that the assumed within-subject working correlation structure is correctly

specified.

Unbalanced longitudinal data are quite common due to missing measurements. When they are missing completely at random, the proposed method can be implemented by transforming the unbalanced data to artificial balanced data as in Zhou and Qu [25]. With $m = \max(m_1, \dots, m_n)$, an $m \times m_i$ transformation matrix M_i is generated by deleting the columns of the $m \times m$ identity matrix corresponding to the missing measurements for subject i . The artificial balanced data is then formulated as $\tilde{X}_i = (I_K \otimes M_i)X_i$ and $\tilde{Y}_i = (I_K \otimes M_i)Y_i$. Accordingly, the estimator can be obtained by maximizing the modified empirical likelihood:

$$\tilde{L}(\beta) = \sup \left\{ \prod_{i=1}^n w_i : \sum_{i=1}^n w_i \tilde{g}_i(\beta) = 0, \sum_{i=1}^n w_i = 1, 0 \leq w_i \leq 1 \right\},$$

where

$$\tilde{g}_i(\beta) = \begin{pmatrix} \tilde{X}_i^\top \tilde{A}_i^{-1/2} \tilde{B}_{i1} \tilde{A}_i^{-1/2} \varphi(\tilde{Y}_i - \tilde{X}_i \beta) \\ \vdots \\ \tilde{X}_i^\top \tilde{A}_i^{-1/2} \tilde{B}_{iq} \tilde{A}_i^{-1/2} \varphi(\tilde{Y}_i - \tilde{X}_i \beta) \end{pmatrix},$$

with $\tilde{A}_i = (I_K \otimes M_i)A_i(I_K \otimes M_i)^\top$ and $\tilde{B}_{ij} = (I_K \otimes M_i)B_{ij}(I_K \otimes M_i)^\top$ for all $j \in \{1, \dots, q\}$. If missing data due to dropouts are caused by a missing at random mechanism, then we can expand the proposed method through an inverse probability weighted estimation procedure as in Lipsitz et al. [13] and Cho et al. [1].

2.2 Inference with multiple quantile regression

An important question with multiple regression quantiles is whether the effect of a certain independent variable varies with the quantiles and exhibits heteroscedasticity of the data. For testing the null hypothesis

$$\mathcal{H}_0 : \beta_{\tau_1, \gamma} = \beta_{\tau_2, \gamma} = \dots = \beta_{\tau_K, \gamma},$$

where for each $k \in \{1, \dots, K\}$, $\beta_{\tau_k, \gamma}$ is the γ th element in the parameter vector β_{τ_k} , the proposed empirical likelihood method readily renders a test: with $\tilde{\beta}$ denoting the maximizer of $L(\beta)$ under the null hypothesis, an empirical likelihood ratio test is constructed with the following test statistic

$$W_n = 2 \ln\{L(\hat{\beta})\} - 2 \ln\{L(\tilde{\beta})\}. \quad (5)$$

The rejection of the null hypothesis suggests the presence of heteroscedasticity in the data. Since the empirical likelihood parallels the parametric counterpart, a nonparametric version of Wilk's theorem holds as stated below.

Theorem 3. *Under regularity conditions 1–3 and the null hypothesis, $W_n \rightsquigarrow \chi_{K-1}^2$ in distribution as $n \rightarrow \infty$, where χ_{K-1}^2 denotes the distribution of a chi-squared random variable with $K - 1$ degrees of freedom.*

With the critical values provided by the limiting distribution $\chi_{1-\alpha, K-1}^2$ at level α , the proposed test is easy to implement. This test could also play an essential role in model diagnostics for other approaches assuming homoscedasticity, e.g., composite quantile regression [7], and optimally weighted quantile average estimation [24]: to this end, we need only modify the null hypothesis as $\mathcal{H}_0 : \beta_{\tau_1, \gamma} = \dots = \beta_{\tau_K, \gamma}$ for all γ except the intercept and test the validity of the homogeneous assumption. The test can also be readily extended to find relevant independent variables with regard to multiple quantiles by setting the appropriate coefficients to zero in the null hypothesis.

The likelihood ratio test can be similarly applied to inference for individual regression coefficients. The literature shows that the empirical likelihood ratio test can be inversely utilized to construct confidence intervals or regions by collecting values for which the test does not reject the null [14]. In practice, however, this inverse method is computationally non-trivial for a multivariate parameter. We instead utilize the asymptotic normality in Theorem 1. Obtaining a consistent plug-in estimator of Σ is challenging, and hence we adopt the random perturbation approach [5]. First we generate n positive random variables V_1, \dots, V_n independently from a distribution with

both mean and variance 1, e.g., the standard unit exponential $\mathcal{E}(1)$. Next we multiply $g_i(\beta)$ by V_i and obtain a perturbed estimator $\hat{\beta}^v$ by maximizing

$$L^v(\beta) = \sup \left\{ \prod_{i=1}^n w_i : \sum_{i=1}^n w_i V_i g_i(\beta) = 0, \sum_{i=1}^n w_i = 1, 0 \leq w_i \leq 1 \right\}.$$

We then evaluate $\hat{\beta}^v$ repeatedly and use the empirical distribution of $\hat{\beta}^v$ to approximate that of $\hat{\beta}$. This resampling procedure is theoretically justified in the Appendix.

3 Simulation studies

We consider the linear model $y_{ij} = \beta_0 + \beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \epsilon_{ij}$ with $i \in \{1, \dots, 100\}$ and $j \in \{1, \dots, 5\}$, where $\beta_0 = \beta_1 = \beta_2 = 1$, $x_{1,ij}$ and $x_{2,ij}$ are generated independently from a $\mathcal{U}(0, 4)$, and the random error $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{i5})^\top$ is generated according to different scenarios as follows:

Case 1: $\epsilon_i \sim \mathcal{N}(0, \Sigma)$, where Σ is an AR(1) correlation matrix with a correlation coefficient of 0.8.

Case 2: $\epsilon_i = e^{\xi_i} - 1$, where $\xi_i \sim \mathcal{N}(0, 0.5\Sigma)$ and Σ is defined in Case 1.

Case 3: $\epsilon_i = (0.7 + 0.2x_{1,i})\eta_i$, where $\eta_i \sim \mathcal{N}(0, \Sigma)$ and Σ is defined in Case 1.

Case 4: $\epsilon_i = e^{\xi_i} + 0.5(x_{1,i} + x_{2,i})\xi_i - 1$, where ξ_i is specified in Case 2.

We assess the conditional quantile regression at $\tau = 0.25, 0.5$ and 0.75 from 1000 simulated data sets. To evaluate the estimation efficiency of the proposed approach, we compute the mean squared error of the estimators based on the multiple quantile regression model under the AR(1) and compound symmetry working correlation structures, and the single quantile regression model under the AR(1), compound symmetry and independent correlations, respectively. Table 1 shows that the proposed approach has much smaller mean squared errors than the one based on the single quantile regression model in all cases under consideration. This illustrates that a substantial efficiency gain is achieved by incorporating the associations across three quantiles. When the choice of working correlation structures is considered, incorporating AR(1) working correlation structure leads to a more efficient estimation compared to ignoring the within-subject correlation and

specifying it incorrectly, while the misspecified structure outperforms the independent correlation structure.

In Cases 3 and 4, heteroscedasticity is present as the random error varies over the independent variables. Thus, we investigate the constancy of the parameter $\beta_{\tau,\gamma}$ at $\tau = 0.25, 0.5$ and 0.75 , for $\gamma = 1$ and 2 individually. The test statistic W_n follows a chi-squared distribution with two degrees of freedom asymptotically if $\beta_{\tau,\gamma}$ is constant at three quartiles. Table 2 summarizes the proportion of times that the null hypothesis is rejected at level of 0.05 out of 1000 simulation runs. In all cases where the constancy null hypothesis is true, the empirical sizes of the test are very close to the nominal level. In the cases where the null hypothesis is not true, the proposed test rejects the hypothesis and detects the differences in the regression quantiles. We also plot empirical quantiles of the test statistics from 1000 simulations under $\mathcal{H}_0 : \beta_{0.25,2} = \beta_{0.5,2} = \beta_{0.75,2}$ against the quantiles of the asymptotic null distributions where the null is true, i.e., in Cases 1, 2 and 3. The results in Figure 1 confirm that the empirical quantiles of W_n follow the theoretical chi-squared quantile fairly well with a modest sample size of 100 .

We further examine the power of the proposed test when the $\beta_{\tau,\gamma}$ s are not constant across quantiles in Cases 3 and 4 by letting $\epsilon_i = (0.7 + cx_{1,i})\eta_i$ and $\epsilon_i = e^{\xi_i} + c(x_{1,i} + x_{2,i})\xi_i - 1$, respectively, where $0 \leq c \leq 1$. We test $\mathcal{H}_0 : \beta_{\tau,\gamma}$ is constant at $\tau = 0.25, 0.5$ and 0.75 for $\gamma = 1$ and 2 . Specifically, we compute W_n under the AR(1) correlation structure for various c from 1000 simulations, and report proportions of times that W_n is greater $\chi_{0.95,2}^2$ in Figure 2. Results in Figure 2 show that when a value of c is 0 , the estimated powers are all close to 0.05 . In the right-hand plot, which corresponds to Case 4, both powers reach 1 when c reaches 0.7 . In Case 3, the power of the test for the constancy of $\beta_{\tau,2}$ also increases as c increases, yet the other power is always around 0.05 regardless of c . This corresponds to the fact that $\beta_{\tau,1}$ is constant regardless of a value of c in Case 3. In sum, the proposed empirical likelihood approach provides not only an easy-to-implement test for detecting change of regression quantiles across different quantiles, but also provides efficient estimators over multiple quantiles.

4 Application to HIV study

In this section, the proposed approach is applied in the analysis of data from a Multi-Center AIDS Cohort study [8]. In general, CD4 cell counts are considered as a biomarker indicating the health status of HIV infected patients. The objective of the study is to evaluate the effects of pre-HIV infection CD4 cell counts, smoking behavior, and the patient's age at HIV infection at three quantile levels ($\tau = 0.25, 0.5$ and 0.75) of CD4 cell counts after the infection over time. This data set consists of 283 homosexual men who became infected by HIV between 1984 and 1991. Although each patient was supposed to be repeatedly measured every 6 months, many patients missed some of their scheduled visits. Moreover, more than half of all patients dropped out after the sixth visit. Thus, we focus on a period of three years when scheduled visits were relatively well kept and formulate the following quantile regression model as

$$Q_{\tau}(y_{ij}|x_{ij}) = \beta_{\tau,0} + \beta_{\tau,1}x_{1,ij} + \beta_{\tau,2}x_{2,i} + \beta_{\tau,3}x_{3,i} + \beta_{\tau,4}x_{4,i},$$

for all $i \in \{1, \dots, 283\}$ and $j \in \{1, \dots, m_i\}$, where $Q_{\tau}(y_{ij}|x_{ij})$ is the conditional τ th quantile of the CD4 cell count given the j th visit time $x_{1,ij}$; $x_{2,i}$ is the centered pre-infection CD4 cell count; $x_{3,i}$ is 1 if the i th subject smoked after infection and 0 otherwise; and $x_{4,i}$ is the centered age at infection. That is, $x_{2,i}$ and $x_{4,i}$ are obtained by subtracting the averages of pre-infection CD4 cell counts and age at infection from the i th subject's pre-infection CD4 cell count and age at infection, respectively. This facilitates the biological interpretation, since $\beta_{\tau,0} + \beta_{\tau,1}x_{1,ij}$ corresponds to the τ th conditional quantile of the CD4 cell count at time $x_{1,ij}$ for a nonsmoker with an average pre-infection CD4 cell count and average age at infection. The other coefficients, $\beta_{\tau,2}$, $\beta_{\tau,3}$, and $\beta_{\tau,4}$, evaluate the effects of the pre-HIV infection CD4 cell count, smoking, and age at HIV-infection on the post-infection CD4 cell count, respectively.

We consider the multiple quantile regression model with $\tau = 0.25, 0.5$ and 0.75 to assess the independent variable effects on the post-CD4 cell count under the AR(1) working correlation struc-

ture. In addition, we estimate standard errors of the estimators via 1000 resampling replications using the random perturbation approach, and assess the statistical significance of the estimated coefficients at a significant level of 0.05. Table 3 reports estimated coefficients with corresponding standard errors. In addition, coefficients whose 95% confidence interval does not include zero are marked with an asterisk (*). The results confirm that the effects of time and pre-infection CD4 cell count were statistically significant with negative and positive estimates respectively for all three quantiles considered. The estimated coefficients also suggest that the true effects of time and pre-infection CD4 cell count vary across quantiles, which may point to heterogeneity in the data.

Thus, we further test $\mathcal{H}_0 : \beta_{0.25,\gamma} = \beta_{0.5,\gamma} = \beta_{0.75,\gamma}$ for $\gamma = 1, \dots, 4$ respectively, and report the test statistics and p -values in Table 3. This confirms that at least one regression quantiles of time and pre-infection CD4 cell counts at $\tau = 0.25, 0.5, 0.75$ is statistically different from the others at a significant level of 0.05. This also corresponds with the results that an association between pre-infection CD4 cell counts and CD4 cell counts after infection at the 0.25th quantile is positive, but weaker than the ones at the median and 0.75th quantile. In addition, over time the CD4 cell counts of patients with initially low CD4 cell counts decrease more rapidly than those of patients with initially high CD4 cell counts as shown in Figure 3.

Appendix: Proofs

Proof of Theorem 1. Recall that we denote the conditional distribution of Y_{ij} given x_{ij} by F_{ij} and the τ th quantile of F_{ij} by $q_{ij}(\tau_k)$, respectively. Let $\beta_{\tau_k 0}$ denote the true τ_k th regression quantile such that $q_{ij}(\tau_k) = x_{ij}^\top \beta_{\tau_k 0}$. Note that under Condition 2, for any $\|\beta_{\tau_k 0} - \beta_k\| = O(n^{-1/2})$, where $\|\cdot\|$ is the Euclidean norm, we have, for each $k \in \{1, \dots, K\}$,

$$\begin{aligned} \mathbb{E}\{\varphi_{\tau_k}(Y_{ij} - x_{ij}^\top \beta_{\tau_k 0}) - \varphi_{\tau_k}(Y_{ij} - x_{ij}^\top \beta_k)\} &= F_{ij}(x_{ij}^\top \beta_{\tau_k}) - F_{ij}(x_{ij}^\top \beta_{\tau_k 0}) \\ &= f_{ij}\{q_{ij}(\tau_k)\}x_{ij}^\top (\beta_k - \beta_{\tau_k 0}) + o(|x_{ij}^\top (\beta_k - \beta_{\tau_k 0})|) \end{aligned}$$

because φ_τ plays a similar role as the indicator function.

As x_{ij} are bounded under Condition 3, by applying Lemma 4.1 of He and Shao [3], we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} -X_i^\top \Delta_i A_i^{-1/2} B_{i1} A_i^{-1/2} X_i \\ \vdots \\ -X_i^\top \Delta_i A_i^{-1/2} B_{iq} A_i^{-1/2} X_i \end{pmatrix} (\beta - \beta_0) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0) + \Gamma(\beta - \beta_0) + o_p(1) \end{aligned} \quad (6)$$

uniformly in β for $\|\beta - \beta_0\| = O(n^{-1/2})$. We let $M_n(\beta) = n^{-1/2} \sum_{i=1}^n g_i(\beta)$ and sketch the main part of the proof below.

It follows from standard empirical likelihood results that

$$\ln L(\beta) = -n^{-1} \sum_{i=1}^n \ln[1 + \{\lambda(\beta)\}^\top g_i(\beta)],$$

where $\lambda(\beta)$ satisfies the equation $n^{-1} \sum_{i=1}^n g_i(\beta) / [1 + \{\lambda(\beta)\}^\top g_i(\beta)] = 0$. Then by Lemma A.4 of Yang and He [22] and (6), we have

$$\begin{aligned} \ln L(\beta) &= - (1/2)(\beta - \beta_0)^\top \Gamma^\top \Omega^{-1} \Gamma (\beta - \beta_0) + n^{-1/2} (\beta - \beta_0)^\top \Gamma^\top \Omega^{-1} \{M_n(\beta_0)\} \\ &\quad - (1/2)n^{-1} \{M_n(\beta_0)\}^\top \Omega^{-1} \{M_n(\beta_0)\} + o_p(n^{-1}) \end{aligned} \quad (7)$$

uniformly in β for $\|\beta - \beta_0\| = O(n^{-1/2})$ and

$$\hat{\beta} - \beta_0 = n^{-1/2} (\Gamma^\top \Omega^{-1} \Gamma)^{-1} \Gamma^\top \Omega^{-1} M_n(\beta_0) + o_p(n^{-1/2}), \quad (8)$$

where Γ and Ω are the same as defined in (4). It follows from $M_n(\beta_0) \rightsquigarrow \mathcal{N}(0, \Omega)$ that $\sqrt{n}(\hat{\beta} - \beta_0) \rightsquigarrow \mathcal{N}(0, \Sigma)$, where $\Sigma = (\Gamma^\top \Omega^{-1} \Gamma)^{-1}$. \diamond

Proof of Theorem 2. The inverse of $R_i(\delta)$ can be represented by $R_i^{-1}(\delta) = G^{-1} \otimes C_i^{-1}(\delta) = I_K \otimes C_i^{-1}(\delta) + W \otimes C_i^{-1}(\delta)$, where W is the matrix containing off-diagonal elements in G^{-1} .

Then, without loss of generality, we pick basis matrices B_{i1}, \dots, B_{iq} such that $I_K \otimes C_i^{-1}(\delta) = \sum_{j=1}^m u_j B_{ij}$ and $W \otimes C_i^{-1}(\delta) = \sum_{j=m+1}^q u_j B_{ij}$, and accordingly define $g_i(\beta)$ as in (3). We further define $Z_{ik} = X_i^\top A_i^{-1/2} B_{ik} A_i^{-1/2}$ for each $k \in \{1, \dots, q\}$. We consider a transformation of Z_{ik} denoted by Z_{ik}^* such that $E\{Z_{ik}^* \varphi(Y_i - X_i \beta_0) \varphi(Y_i - X_i \beta_0)^\top Z_{i\ell}^{*\top}\} = 0$ for any $1 \leq k \leq m$ and $(m+1) \leq \ell \leq q$, and $E\{Z_{ik}^* \varphi(Y_i - X_i \beta_0) \varphi(Y_i - X_i \beta_0)^\top Z_{i\ell}^{*\top}\} = 0$ for any $(m+1) \leq k < \ell \leq q$. Then,

$$\Sigma^{-1} = \Sigma_C^{-1} + \sum_{j=(m+1)}^q E(X_i^\top \Delta_i Z_{ij}^{*\top}) E\{Z_{ij}^* \varphi(Y_i - X_i \beta_0) \varphi(Y_i - X_i \beta_0)^\top Z_{ij}^{*\top}\} E(Z_{ij}^* \Delta_i X_i).$$

The proof is completed by the fact that for each $j \in \{m+1, \dots, q\}$, $E\{Z_{ij}^* \varphi(Y_i - X_i \beta_0) \varphi(Y_i - X_i \beta_0)^\top Z_{ij}^{*\top}\}$ is positive definite since $E\{g_i(\beta_0) g_i(\beta_0)^\top\}$ is positive definite under Condition 1. In a similar manner as above, it can be readily proved that $\Sigma_I - \Sigma_{C_i}$ is semi-positive definite. \diamond

Proof of Theorem 3. Recall the null hypothesis $\mathcal{H}_0 : \beta_{\tau_1, \gamma} = \beta_{\tau_2, \gamma} = \dots = \beta_{\tau_K, \gamma}$. We consider $K - 1$ contrasts $\beta_{\tau_j, \gamma} - \beta_{\tau_K, \gamma}$, where $j \in \{1, \dots, K - 1\}$ and note that the null hypothesis is equivalent to constraining $\beta_{\tau_j, \gamma} - \beta_{\tau_K, \gamma} = 0$ for all $j \in \{1, \dots, K - 1\}$. We consider a linear transformation of $g_i(\beta)$ defined by a square matrix H^* such that the first $K - 1$ elements correspond to the contrasts and the rest elements not constrained are identical to $g_i(\beta)$. For notational simplicity, we denote the parameters on the transformed scale by β^* , and let $\beta^* = (\beta_1^*, \beta_2^*)$ with β_1 corresponding to the $K - 1$ contrasts. We accordingly denote Γ and Ω on the transformed scale by Γ^* and Ω^* , where $\Gamma^* = (\Gamma_1^*, \Gamma_2^*)$ corresponds to $\beta^* = (\beta_1^*, \beta_2^*)$.

Then we note that $W_n = 2 \ln\{L(\hat{\beta}_1^*, \hat{\beta}_2^*)\} - 2 \ln\{L(0, \tilde{\beta}_2^*)\}$. With (7) and (8), it follows from similar lines of argument as in the proof of Corollary 5 of Qin and Lawless [16] that

$$\begin{aligned} W_n &= 2 \ln\{L(\hat{\beta}_1^*, \hat{\beta}_2^*)\} - 2 \ln\{L(0, \tilde{\beta}_2^*)\} \\ &= M_n^*(\beta_0^*)^\top \Omega^{*-1} [\Gamma^* \{\Gamma^{*\top} \Omega^{*-1} \Gamma^*\}^{-1} \Gamma^{*\top} - \Gamma_1^* \{\Gamma_1^{*\top} \Omega^{*-1} \Gamma_1^*\}^{-1} \Gamma_1^{*\top}] \Omega^{*-1} M_n^*(\beta_0^*) + o_p(1), \end{aligned}$$

where $\beta_0^* = (0, \beta_{20}^*)$ with β_{20}^* denoting the true value of β_2^* and

$$\begin{aligned} \Gamma^* \{\Gamma^{*\top} \Omega^{*-1} \Gamma^*\}^{-1} \Gamma^{*\top} &\geq \begin{pmatrix} \Gamma_1^* & \Gamma_2^* \end{pmatrix} \begin{pmatrix} \{\Gamma_1^{*\top} \Omega^{*-1} \Gamma_1^*\}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Gamma_1^{*\top} \\ \Gamma_2^{*\top} \end{pmatrix} \\ &= \Gamma_1^* \{\Gamma_1^{*\top} \Omega^{*-1} \Gamma_1^*\}^{-1} \Gamma_1^{*\top}. \end{aligned}$$

As $\Gamma^* \{\Gamma^{*\top} \Omega^{*-1} \Gamma^*\}^{-1} \Gamma^{*\top}$ and $\Gamma_1^* \{\Gamma_1^{*\top} \Omega^{*-1} \Gamma_1^*\}^{-1} \Gamma_1^{*\top}$ are symmetric idempotent matrices with trace respectively equal to $pKq - (pKq - pK)$ and $pK - (K - 1)$ similarly by the Corollary 5 of Qin and Lawless [16], one can conclude that $W_n \rightsquigarrow \chi_{K-1}^2$ as $n \rightarrow \infty$. \diamond

Proof of the resampling procedure. Let $M_n^v(\beta) = n^{-1/2} \sum_{i=1}^n V_i g_i(\beta)$. Then, similar to (6), we can obtain $M_n^v(\beta) = M_n^v(\hat{\beta}) + \Gamma(\beta - \hat{\beta}) + o_p(1)$ uniformly in β for $\|\beta - \hat{\beta}\| = O(n^{-1/2})$. Following Proposition A.3 in Jin et al. [5] and repeating the steps from (6) to (8), we have that, conditioning on the data,

$$\Omega^v = E\{M_n^v(\hat{\beta}) M_n^v(\hat{\beta})^\top\} = \frac{1}{n} \sum_{i=1}^n E\{V_i^2 g_i(\hat{\beta}) g_i(\hat{\beta})^\top\} = \frac{1}{n} \sum_{i=1}^n E\{g_i(\hat{\beta}) g_i(\hat{\beta})^\top\} \rightarrow \Omega$$

as $n \rightarrow \infty$ and

$$\sqrt{n}(\hat{\beta}^v - \hat{\beta}) = (\Gamma^\top \Omega^{-1} \Gamma)^{-1} \Gamma^\top \Omega^{-1} M_n^v(\hat{\beta}) + o_p(n^{-1/2}).$$

Consequently, conditioning on the data, since $M_n^v(\hat{\beta}) \rightsquigarrow N(0, \Omega)$ by Slutsky's Lemma, we have

$$\sqrt{n}(\hat{\beta}^v - \hat{\beta}) \rightsquigarrow \mathcal{N}[0, (\Gamma^\top \Omega^{-1} \Gamma)^{-1}].$$

This concludes the argument. \diamond

References

- [1] Cho, H., Hong, H., and Kim, M. (2016). Efficient quantile marginal regression for longitudinal data with dropouts. *Biostatistics* **17**, 561–575.
- [2] Furno, M. (2004). ARCH tests and quantile regressions. *Journal of Statistical Computation & Simulation* **74**, 277–292.
- [3] He, X. and Shao, Q.M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics* **24**, 2608–2630.
- [4] He, X., Fu, B., and Fung, W.K. (2003). Median regression for longitudinal data. *Statistics in Medicine* **22**, 3655–3669.
- [5] Jin, Z., Ying, Z., and Wei, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–390.
- [6] Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91**, 251–257.
- [7] Kai, B., Li, R., and Zou, H. (2010). Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society: Series B* **72**, 49–69.
- [8] Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F., and Rinaldo, C.R. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristics of the participants. *American Journal of Epidemiology* **126**, 310–318.
- [9] Koenker, R. (1984). A note on L-estimators for linear models. *Statistics and Probability Letters* **2**, 323–325.

- [10] Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91**, 74–89.
- [11] Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantile. *Econometrica* **50**, 43–61.
- [12] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalised linear models. *Biometrika* **73**, 12–22.
- [13] Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., and Zhao, L.P. (1997). Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Applied Statistics* **46**, 463–476.
- [14] Owen, A.B. (2001). *Empirical likelihood*. New York: Chapman and Hall-CRC.
- [15] Portnoy, S. and Koenker, R. (1989). Adaptive L-estimation of linear models. *The Annals of Statistics* **17**, 362–381.
- [16] Qin, J. and Lawless, J.F. (1994). Empirical likelihood and generalized estimating equations. *The Annals of Statistics* **22**, 300–25.
- [17] Qu, A., Lindsay, B.G., and Li, B. (2000). Improving estimating equations using quadratic inference functions. *Biometrika* **87**, 823–36.
- [18] Tang, C.Y. and Leng, C. (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika* **98**, 1001–1006.
- [19] Tang, Y., Wang, Y., Li, J., and Qian, W. (2015). Improving estimation efficiency in quantile regression with longitudinal data. *Journal of Statistical Planning and Inference* **165**, 38–55.
- [20] Wang, H. and Zhu, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *Journal of Statistical Planning and Inference* **141**, 1603–1615.

- [21] Xiao, Z. and Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* **104**, 1696–1712.
- [22] Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression, *The Annals of Statistics* **40**, 1102–1131.
- [23] Yi, G. and He, W. (2009). Median regression models for longitudinal data with dropouts. *Biometrics* **65**, 618–625.
- [24] Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory* **30**, 1272–1314.
- [25] Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistics Association* **107**, 701–710.
- [26] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108–1126.

Table 1: Mean squared errors ($\times 1000$) of estimators based on the multiple quantile regression model under the AR(1) correlation structure (Multi_{AR}) and the compound symmetry structure (Multi_{CS}), and the single quantile regression model under the AR(1) (Single_{AR}), the compound symmetry (Single_{CS}) and independent correlation structures (Single_{IN}).

	$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
<i>Case 1</i>									
Multi _{AR}	2.93	0.50	0.39	1.79	0.35	0.35	2.85	0.63	0.66
Multi _{CS}	3.19	0.51	0.46	1.79	0.38	0.40	3.17	0.64	0.85
Single _{AR}	7.51	1.66	1.56	6.44	1.23	1.15	11.95	1.50	1.51
Single _{CS}	8.41	1.71	1.66	9.03	1.26	1.30	15.94	1.58	1.57
Single _{IN}	12.71	2.28	2.32	15.44	2.07	1.93	19.64	2.53	2.35
<i>Case 2</i>									
Multi _{AR}	0.55	0.06	0.06	0.63	0.10	0.10	1.95	0.40	0.41
Multi _{CS}	0.59	0.08	0.08	0.65	0.11	0.12	1.96	0.42	0.44
Single _{AR}	2.51	0.19	0.18	3.24	0.32	0.34	6.93	0.83	0.78
Single _{CS}	2.53	0.20	0.20	3.82	0.34	0.38	7.63	0.92	0.81
Single _{IN}	3.10	0.29	0.30	5.15	0.53	0.59	10.70	1.26	1.44
<i>Case 3</i>									
Multi _{AR}	2.56	0.87	0.45	1.44	0.56	0.28	2.47	1.10	0.66
Multi _{CS}	2.52	0.88	0.49	1.45	0.58	0.36	2.50	1.23	0.70
Single _{AR}	6.98	2.41	1.37	7.06	1.93	1.23	11.14	2.55	1.59
Single _{CS}	9.13	2.65	1.60	8.28	2.04	1.32	12.49	2.89	1.57
Single _{IN}	14.15	3.65	2.39	13.79	2.86	2.13	18.99	3.35	2.36
<i>Case 4</i>									
Multi _{AR}	2.20	1.11	1.24	1.69	0.81	0.78	3.13	2.11	1.71
Multi _{CS}	2.67	1.27	1.25	1.87	0.84	0.78	3.15	2.15	1.85
Single _{AR}	7.28	3.25	3.21	7.85	3.22	2.97	15.35	5.23	4.74
Single _{CS}	8.20	3.30	3.33	10.77	3.30	3.27	17.09	5.25	4.80
Single _{IN}	14.24	4.44	4.37	16.86	4.83	4.50	27.83	6.78	7.09

Table 2: Proportions of times that we reject $\mathcal{H}_0 : \beta_{\tau,\gamma}$ is constant at $\tau = 0.25, 0.5$ and 0.75 for $\gamma = 1$ and 2 .

	<i>Case 1</i>	<i>Case 2</i>	<i>Case 3</i>	<i>Case 4</i>
$\beta_{\tau,1}$	0.041	0.044	0.962	0.956
$\beta_{\tau,2}$	0.045	0.036	0.039	0.949

Table 3: Estimated coefficients with the standard errors and test statistics with the p -values for heteroscedasticity of data; a coefficient whose 95% confidence interval does not include zero is marked with an asterisk (*)

τ	Intercept _{se}	Time _{se}	PreCD4 _{se}	Smoke _{se}	Age _{se}
0.25	29.04 _{0.81} *	-3.41 _{0.34} *	0.28 _{0.09} *	1.92 _{1.33}	-0.10 _{0.11}
0.5	34.72 _{1.03} *	-2.29 _{0.38} *	0.48 _{0.10} *	1.06 _{1.45}	-0.12 _{0.09}
0.75	41.33 _{0.87} *	-1.89 _{0.38} *	0.52 _{0.08} *	0.14 _{1.24}	-0.13 _{0.11}

Heteroscedasticity test				
Test statistic	7.60	11.86	3.22	1.49
p -value	0.022	0.003	0.200	0.475

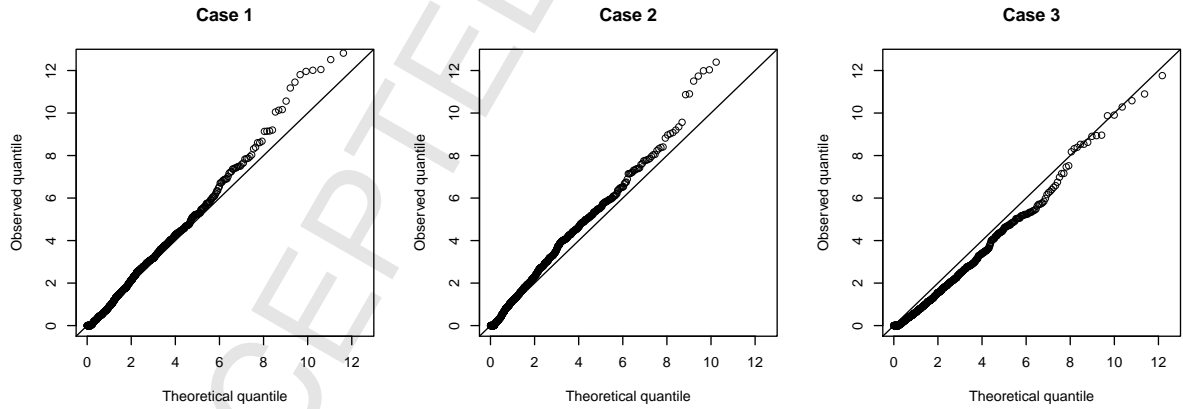


Figure 1: Quantile-quantile plots for the chi-square distribution χ^2_2 versus the test statistic W_n using the proposed approach under \mathcal{H}_0 : $\beta_{\tau,2}$ is constant across three different quantiles in Cases 1, 2 and 3.

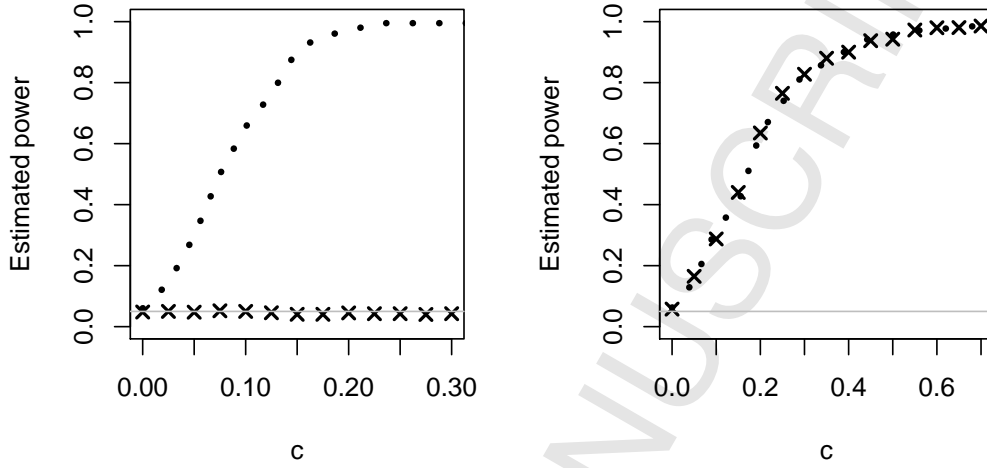


Figure 2: Estimated power curves of the empirical likelihood ratio tests using the proposed approach under $\mathcal{H}_0 : \beta_{0.25,\gamma} = \beta_{0.5,\gamma} = \beta_{0.75,\gamma}$ for $\gamma = 1$ and 2 in Cases 3 (left) and 4 (right). The dotted curves and x-marked curves are estimated powers at $\gamma = 1$ and 2 , respectively, and gray lines indicate the nominal level of 0.05 .

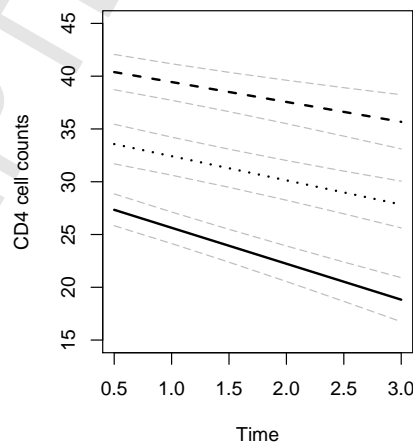


Figure 3: Fitted CD4 cell counts based on the proposed approach at $\tau = 0.25$ (solid line), 0.50 (dotted line), and 0.75 (black dashed line) along with 95% confidence intervals (grey dashed lines).