Weighted quantile regression for longitudinal data

Xiaoming Lu, Zhaozhi Fan

Department of Mathematics and Statistics, Memorial University of Newfoundland, St.John's, Newfoundland, Canada

Abstract

Quantile regression is a powerful statistical methodology that complements the classical linear regression by examining how covariates influence the location, scale, and shape of the entire response distribution and offering a global view of the statistical landscape. In this paper we propose a new quantile regression model for longitudinal data. The proposed approach incorporates the correlation structure between repeated measures to enhance the efficiency of the inference. In order to use the Newton-Raphson iteration method to obtain convergent estimates, the estimating functions are redefined as smoothed functions which are differentiable with respect to regression parameters. Our proposed method for quantile regression provides consistent estimates with asymptotically normal distributions. Simulation studies are carried out to evaluate the performance of the proposed method. As an illustration, the proposed method was applied to a real-life data that contains self-reported labor pain for women in two groups.

Keywords: Quantile regression, Longitudinal data, Quasi-likelihood, Correlation

1. Introduction

Longitudinal data are very common in many areas of applied studies. Such data are repeatedly collected from independent subjects over time and correlation arises between measures from the same subject. One advantage of longitudinal study is that, additional to modeling the cohort effect, one can still specify the individual patterns of change. In order to take the correlation into consideration to not only avoid loss of efficiency in estimation but also make correct statistical inference, a number of methods are developed to evaluate covariate effects on the mean of a response variable (Liang and Zeger, 1986; Qu et al., 2000; Jung and Ying, 2003). Sutradhar (2003) has proposed a generalization of the quasi-likelihood estimation approach to model the conditional mean of the response by solving the generalized quasi-likelihood (GQL) estimating equations. A general stationary auto-correlation matrix is used in this method, which, in fact, represents the correlations of many stationary dynamic models, such as stationary auto-regressive order 1 (AR(1)), stationary moving average order 1 (MA(1)), and stationary equi-correlation (EQC) models.

Quantile regression (Koenker and Bassett Jr, 1978) has become a widely used technique in applications. The effects of covariates are modeled through conditional quantiles of the response variable, rather than the conditional mean, which makes it possible to characterize any arbitrary point of a distribution and thus provide a complete description of the entire response distribution. Compared to the classical mean regression, quantile regression is more robust to outliers and the error patterns do not need to be specified. Therefore, quantile regression has been widely used, (see Chen et al., 2004; Koenker, 2005; Reich et al., 2010; Farcomeni, 2012, among others).

Recently quantile regression has been extended to longitudinal data analysis. A simple way to do so is to assume working independence that ignores correlations between repeated measures, which, of course, may cause loss of efficiency, see Wei and He (2006); Wang and He (2007); Mu and Wei (2009); Wang and Fygenson (2009); Wang (2009); Wang et al. (2009). Jung (1996) firstly developed a quasi-likelihood method for median regression which incorporates correlations between repeated measures for longitudinal data. This method requires estimation of the correlation matrix. Based on Jung's work, Lipsitz et al. (1997) proposed a weighted GEE model. Koenker (2004) considered a random effect model for estimating quantile functions with subject specific fixed effects and based inference on a penalized likelihood. Karlsson (2008) suggested a weighted approach for a nonlinear quantile regression estimation of longitudinal data. Geraci and Bottai (2007) made inferences by using a random intercept to account for the within-subject correlations and proposed a method using the asymmetric Laplace distribution (ALD). Geraci and Bottai's work was generalized by Liu and Bottai (2009), who gave a linear mixed effect quantile regression model using a multivariate Laplace distribution. Farcomeni (2012) proposed a linear quantile regression model allowing time-varying random effects and modeled subject-specific parameters through a latent Markov chain. To reduce the loss of efficiency in inferences of quantile regression, Tang and Leng (2011) incorporate the within-subject correlations through a specified conditional mean model.

Unlike in the classical linear regression, it is difficult to account for the correlations between repeated measures in quantile regression. Misspecification of the correlation structure in GEE method also leads to loss of inferential efficiency. Moreover, the approximating algorithms for computing estimates could be very complicated, and computational problems could occur when statistical software is applied to do intensive re-samplings in the inference procedure. To overcome these problems, Fu and Wang (2012) proposed a combination of the between- and within-subject estimating equations for parameter estimation. By combining multiple sets of estimating equations, Leng and Zhang (2012) developed a new quantile regression model which produces efficient estimates. Those two papers extend the induced smoothing method (Brown and Wang, 2005) to quantile regression, and thus obtained smoothed objective functions which allow the application of Newton-Raphson iteration, and the latter automatically gives both the estimates of parameters and the sandwich estimate of their covariance matrix.

In this paper, we propose a more general quantile regression model by appropriately incorporating a correlation structure between repeated measures in longitudinal data. By employing a general stationary auto-correlation matrix, we avoid the specification of any particular correlation structure. The correlation coefficients can be iteratively estimated in the process of the regression estimation. By using the induced smoothed estimating functions, we can obtain estimates of parameters and their asymptotic covariance matrix by using Newton-Raphson algorithm. The estimators obtained using our proposed method are consistent and asymptotically normal. The results of the intensive simulation studies reveal that our proposed method outperforms those methods based on working independence assumption. Furthermore, our approach is simpler and more general than other quantile regression methods for longitudinal data on theoretical derivation, practical application and statistical programming.

The remainder of this paper proceeds as follows: In Section 2 we develop the proposed quantile regression method and the algorithm of parameter estimation. The asymptotic properties of the proposed estimators are discussed in Section 3. Intensive simulation studies were carried out and results are presented in Section 4. Application of our method to the labor pain data is presented in Section 5. The paper is concluded in Section 6 with some concluding remarks.

2. Proposed quantile regression models

Suppose, in a longitudinal setup, we collect a small number of repeated responses along with certain multidimensional covariates from a large number of independent individuals. Let $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ denote $n_i \geq 2$ repeated measures observed from the ith subject, for $i = 1, \ldots, m$ where m is a positive integer. Let $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ be the p-dimensional covariate vector corresponding to y_{ij}. Suppose that responses from different individuals are independent and those from the same subject are dependent. Let the conditional τth quantile of y_{ij} given x_{ij} be

$$Q_\tau(y_{ij}|x_{ij}) = x_{ij}^T \beta_\tau.$$

In quantile regression we are interested in estimating β_τ consistently and as efficiently as possible.

If we assume the working independence (WI) between repeated measures of responses among each individual, we can obtain $\hat{\beta}_{WI\tau}$, an estimate of β_τ, by minimizing the following objective function

$$S(\beta_\tau) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \rho_\tau(y_{ij} - x_{ij}^T \beta_\tau), \tag{1}$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ (Koenker and Bassett Jr, 1978). Estimating equations can be derived from function (1) by equating the differentiation of $S(\beta_\tau)$ with respect to β_τ to 0. That is

$$U_0(\beta_\tau) = \frac{\partial S(\beta_\tau)}{\partial \beta_\tau} = \sum_{i=1}^{m} X_i^T \psi_\tau(y_i - X_i \beta_\tau) = 0, \tag{2}$$

where $\psi_\tau(u) = \rho'_\tau(u) = \tau - I(u < 0)$ is a discontinuous function, and $\psi_\tau(y_i - X_i\beta_\tau) = (\psi_\tau(y_{i1} - x_{i1}^T\beta_\tau), \ldots, \psi_\tau(y_{in_i} - x_{in_i}^T\beta_\tau))^T$ is a $n_i \times 1$ vector. An efficient algorithm to obtain an estimate of β_τ by solving the equation (2), $U_0(\beta_\tau) = 0$, was given by Koenker and D'Orey (1987), which is available in statistical software R (package quantreg). Parameter estimator $\hat{\beta}_{WI\tau}$ is derived from estimating equation (2) under working independence assumption, therefore the efficiency of $\hat{\beta}_{WI\tau}$ may not be satisfactory.

To take the within correlations into consideration when constructing quantile regression models for longitudinal data, a quasi-likelihood (QL) method was introduced by Jung (1996). Let $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{ij}, \ldots, \varepsilon_{in_i})^T$, where $\varepsilon_{ij} = y_{ij} - x_{ij}^T \beta_\tau$ which is a continuous error term satisfying $P(\varepsilon_{ij} \leq 0) = \tau$ and with an unknown density function $f_{ij}(\cdot)$. In least squares, or mean regression model, Bernoulli distributed $\psi_\tau(\varepsilon_i) = \psi_\tau(y_i - X_i\beta_\tau)$ can be treated as a random noise vector. Using this fact, the QL can be generalized into quantile regression by estimating the correlation matrix of $\psi_\tau(\varepsilon_i)$. Let the covariance matrix of $\psi_\tau(\varepsilon_i)$ be denoted as

$$V_i = \mathrm{cov}(\psi_\tau(y_i - X_i\beta_\tau)) = \mathrm{cov}\begin{pmatrix} \tau - I(\varepsilon_{i1} < 0) \\ \vdots \\ \tau - I(\varepsilon_{in_i} < 0) \end{pmatrix},$$

and

$$\varGamma_i = \mathrm{diag}[f_{i1}(0), \ldots, f_{in_i}(0)] = \begin{pmatrix} f_{i1}(0) & & \\ & \ddots & \\ & & f_{in_i}(0) \end{pmatrix},$$

be an $n_i \times n_i$ diagonal matrix with jth diagonal element $f_{ij}(0)$. Jung (1996) derived the derivative of the log-quasi-likelihood $l(\beta_\tau; y_i)$ with respect to β_τ, which can be used to estimate β_τ by solving

$$U_1(\beta_\tau) = \sum_{i=1}^{m} \frac{\partial l(\beta_\tau; y_i)}{\partial \beta_\tau} = \sum_{i=1}^{m} X_i^T \varGamma_i V_i^{-1} \psi_\tau(y_i - X_i\beta_\tau) = 0. \tag{3}$$

In estimating equation (3), the term $\varGamma_i$ describes the dispersions in ε_{ij} and its diagonal elements can be well estimated by following Hendricks and Koenker (1992):

$$\hat{f}_{ij}(0) = 2h_n[x_{ij}^T(\hat{\beta}_{\tau+h_n} - \hat{\beta}_{\tau-h_n})]^{-1},$$

where $h_n \to 0$ when $n \to \infty$ is a bandwidth parameter. In some cases when f_{ij} is difficult to estimate, $\varGamma_i$ can be simply treated as an identity matrix with a slight loss of efficiency (Jung, 1996).

However, the estimation of the covariance matrix V_i becomes much complicated when QL method is applied. Whatever correlation matrix that ε_i follows, the correlation matrix of $\psi_\tau(\varepsilon_i)$ is no longer the same one, and its correlation structure may be very difficult to specify.

Here, we propose a new method based on the following estimating equations

$$U(\beta_\tau) = \sum_{i=1}^{m} X_i^T \varGamma_i \varSigma_i^{-1}(\rho) \psi_\tau(y_i - X_i\beta_\tau) = 0, \tag{4}$$

where $\varSigma_i(\rho)$ is the covariance matrix of $\psi_\tau(\varepsilon_i)$ that can be expressed as $\varSigma_i(\rho) = A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}$, with $A_i = \mathrm{diag}[\sigma_{i11}, \ldots, \sigma_{1n_in_i}]$ being an $n_i \times n_i$ diagonal matrix, $\sigma_{ijj} = \mathrm{var}(\psi_\tau(\varepsilon_{ij}))$ and $C_i(\rho)$ as the correlation matrix of $\psi_\tau(\varepsilon_i)$, ρ being a correlation index parameter. Suppose the matrix $\Sigma_i(\rho)$ in equation (4) has a general stationary auto-correlation structure such that the correlation matrix $C_i(\rho)$ takes the form of

$$C_i(\rho) = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n_i-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n_i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{n_i-1} & \rho_{n_i-2} & \rho_{n_i-3} & \cdots & 1 \end{pmatrix}$$

for all $i = 1, \ldots, m$, where ρ_ℓ can be estimated by

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i-\ell} \tilde{y}_{ij}\tilde{y}_{i,j+\ell}/m(n_i-\ell)}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} \tilde{y}_{ij}^2/mn_i}$$

for $\ell = 1, \ldots, n_i - 1$ (Sutradhar and Kovacevic, 2000; Sutradhar, 2003) with $\tilde{y}_{ij}$ defined as

$$\tilde{y}_{ij} = \frac{\psi_\tau(y_{ij} - x_{ij}^T\beta_\tau)}{\sqrt{\sigma_{ijj}}}.$$

To estimate $\sigma_{ijj} = \mathrm{var}(\psi_\tau(y_{ij} - x_{ij}^T\beta_\tau))$, we apply the fact that $\psi_\tau(\varepsilon_{ij}) = \psi_\tau(y_{ij} - x_{ij}^T\beta_\tau) = \tau - I(y_{ij} < x_{ij}^T\beta_\tau)$. Hence we have

$$\begin{aligned}\sigma_{ijj} &= \mathrm{var}[\psi_\tau(\varepsilon_{ij})] = \mathrm{var}[\tau - I(y_{ij} < x_{ij}^T\beta_\tau)] = \mathrm{var}[I(y_{ij} < x_{ij}^T\beta_\tau)] \\ &= \Pr(y_{ij} < x_{ij}^T\beta_\tau)(1 - \Pr(y_{ij} < x_{ij}^T\beta_\tau)),\end{aligned}$$

where $\Pr(y_{ij} < x_{ij}^T\beta_\tau)$ is the probability of the event $\{y_{ij} < x_{ij}^T\beta_\tau\}$. If β_τ is the true parameter, we know that $x_{ij}^T\beta_\tau$ is exactly the τth quantile of the variable y_{ij}, hence $\Pr(y_{ij} < x_{ij}^T\beta_\tau) = \tau$, which leads to an estimator of σ_{ijj}, $\tilde{\sigma}_{ijj} = \tau(1-\tau)$. Consequently, A_i matrix can be calculated at the true β_τ as

$$\begin{aligned}\tilde{A}_i &= \mathrm{diag}[\tilde{\sigma}_{i11}, \ldots, \tilde{\sigma}_{1n_in_i}] \\ &= \begin{pmatrix} \tau(1-\tau) & & \\ & \ddots & \\ & & \tau(1-\tau) \end{pmatrix}_{n_i \times n_i},\end{aligned} \tag{5}$$

indicating a constant diagonal matrix for a certain τ. We denote the parameter estimator obtained from this proposed quantile regression (PQR) model as $\hat{\beta}_{PQR\tau}$.

Notice that in the expression $\Sigma_i(\rho) = A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}$, if we set A_i as the one at the true β_τ which is given by (5), $C_i(\rho)$ becomes the only part in $\Sigma_i(\rho)$ containing the information about the data and the parameter β_τ. However in practice, the estimated parameter may never be exactly the true β_τ. Thus, the elements of the diagonal matrix A_i may differ from the constant value $\tau(1-\tau)$. Moreover, we expect the matrix A_i to be also related to the parameter estimates, which becomes crucial when we use an iteration method to estimate parameters where the estimates $\hat{\beta}_\tau$ need to be updated within each iteration step. In this case, as long as the sample size is large enough, we can estimate the diagonal elements of A_i by the following

$$\begin{aligned}\hat{\sigma}_{ijj} &= \Pr(y_{ij} < x_{ij}^T\beta_\tau)(1 - \Pr(y_{ij} < x_{ij}^T\beta_\tau)) \\ &= \frac{1}{m}\sum_{i=1}^{m} I(y_{ij} < x_{ij}^T\beta_\tau)(1 - \frac{1}{m}\sum_{i=1}^{m} I(y_{ij} < x_{ij}^T\beta_\tau)),\end{aligned}$$

for all $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$. By using $\hat{\sigma}_{ijj}$ to estimate Σ_i, the solution-finding iteration converges faster. The solution of estimating equations (4) leads to an adjusted estimate of β_τ. We call this method as adjusted quantile regression (AQR).

The difficulty of solving the estimating equation (4) is caused by the non-convex and non-continuous objective function $U(\beta_\tau)$ which is not differentiable. Though several methods can be applied to estimate β_τ from equation (4) without requiring any derivatives and continuity of the estimating function, they may become very complicated and cause a high burden of computation. To overcome these difficulties, the induced smoothing method has been extended to the quantile regression for longitudinal data assuming a working correlation by Fu and Wang (2012). The smoothing method is asymptotically equivalent to its original counterpart, see Lemma 3.1 below. Here, let $\tilde{U}(\beta_\tau) = E_Z[U(\beta_\tau + \Omega^{1/2}Z)]$, with expectation taken with respect to Z, where $Z \sim N(0, I_p)$, and Ω is updated as an estimate of the covariance matrix of parameter estimators. After some algebraic calculations, a smoothed estimating function $\tilde{U}(\beta_\tau)$ is obtained as

$$\tilde{U}(\beta_\tau) = \sum_{i=1}^{m} X_i^T \Gamma_i \Sigma_i^{-1}(\rho)\tilde{\psi}_\tau(y_i - X_i\beta_\tau) \tag{6}$$

where

$$\tilde{\psi}_\tau = \begin{pmatrix} \tau - 1 + \Phi(\frac{y_{i1} - x_{i1}^T\beta_\tau}{r_{i1}}) \\ \vdots \\ \tau - 1 + \Phi(\frac{y_{in_i} - x_{in_i}^T\beta_\tau}{r_{in_i}}) \end{pmatrix}$$

and $r_{ij} = \sqrt{x_{ij}^T\Omega x_{ij}}$ for $j = 1, \ldots, n_i$. Thus the differentiation of $\tilde{U}(\beta_\tau)$ with respect to β_τ can be easily calculated, and we can use $\partial\tilde{U}(\beta_\tau)/\partial\beta_\tau$ as an approximation of $\partial U(\beta_\tau)/\partial\beta_\tau$ as

$$\frac{\partial\tilde{U}(\beta_\tau)}{\partial\beta_\tau} = -\sum_{i=1}^{m} X_i^T\Gamma_i\Sigma_i^{-1}(\rho)\tilde{\Lambda}_i X_i,$$

where $\tilde{\Lambda}_i$ is an $n_i \times n_i$ diagonal matrix with the jth diagonal element $\phi((y_{ij} - x_{ij}^T\beta_\tau)/r_{ij})/r_{ij}$. Generally, let $\hat{\beta}_{WI\tau}$ be the estimate under the working independence assumption and I_p be a identity matrix of size p, smoothed estimators of β_τ and its covariance matrix Ω can be obtained from the following Newton-Raphson iteration:

Step 1. Given initial values of β_τ and the symmetric positive definite matrix Ω as $\tilde{\beta}_\tau(0) = \hat{\beta}_{WI\tau}$ and $\tilde{\Omega}(0) = \frac{1}{m}I_p$ respectively.

Step 2. Using $\tilde{\beta}_\tau(r)$ and $\tilde{\Omega}(r)$ given from the rth iteration, update $\tilde{\beta}_\tau(r+1)$ and $\tilde{\Omega}(r+1)$ by

$$\tilde{\beta}_\tau(r+1) = \tilde{\beta}_\tau(r) + \left[-\frac{\partial \tilde{U}(\beta_\tau)}{\partial \beta_\tau}\right]_r^{-1} \times \left[\tilde{U}(\beta_\tau)\right]_r \quad \text{and}$$

$$\tilde{\Omega}(r+1) = \left[-\frac{\partial \tilde{U}(\beta_\tau)}{\partial \beta_\tau}\right]_r^{-1} \times \left[\operatorname{cov}(\tilde{U}(\beta_\tau))\right]_r \times \left[-\frac{\partial \tilde{U}(\beta_\tau)}{\partial \beta_\tau}\right]_r^{-1},$$

where $[]_r$ denotes that the expression between the square brackets is evaluated at $\beta_\tau = \tilde{\beta}_\tau(r)$ and $\operatorname{cov}(\tilde{U}(\beta_\tau)) = \sum_{i=1}^m X_i^T \Gamma_i \Sigma_i^{-1}(\rho) \tilde{\psi}_\tau(\varepsilon_i) \tilde{\psi}_\tau^T(\varepsilon_i) \Sigma_i^{-1}(\rho) \Gamma_i X_i$.

Step 3. Repeat step 2 until convergence.

This method provides consistent estimates of β_τ and its covariance matrix Ω. Furthermore, compared with other techniques, our method based on Newton-Raphson algorithm is much faster.

3. Asymptotic properties

In this section, we derive asymptotic distributions of the proposed estimates obtained from both the original estimating equation (4) and smoothed estimating equation (6).

Theorem 3.1. *Under regularity conditions A1-A5 listed in Appendix, the estimator $\hat{\beta}_\tau$ based on the original estimating equation (4) is $\sqrt{m}$-consistent and asymptotically normal,*

$$\sqrt{m}(\hat{\beta}_\tau - \beta_\tau) \to N(0, G^{-1}(\beta_\tau) V \{G^{-1}(\beta_\tau)\}^T),$$

where, in the variance-covariance matrix, $G(\beta_\tau) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^m X_i^T \Gamma_i \Sigma_i^{-1}(\rho) \Gamma_i X_i$ and $V = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^m X_i^T \Gamma_i \Sigma_i^{-1}(\rho) \operatorname{cov}\{\psi_\tau(y_i - X_i\beta_\tau)\} \Sigma_i^{-1}(\rho) \Gamma_i X_i$.

Lemma 3.1. *Under regularity conditions A1-A5 listed in Appendix, the smoothed estimating functions $\tilde{U}(\beta_\tau)$ are asymptotically equivalent to the original estimating functions $U(\beta_\tau)$ in the sense that,*

$$\frac{1}{\sqrt{m}}\{\tilde{U}(\beta_\tau) - U(\beta_\tau)\} = o_p(1).$$

Theorem 3.2. *Under regularity conditions A1-A5 listed in Appendix, the estimator $\tilde{\beta}_\tau$ based on the smoothed estimating equation (4) is $\sqrt{m}$-consistent and asymptotically normal,*

$$\sqrt{m}(\tilde{\beta}_\tau - \beta_\tau) \to N(0, G^{-1}(\beta_\tau) V \{G^{-1}(\beta_\tau)\}^T),$$

where $G(\beta_\tau)$ and V have the same expressions as in Theorem 3.1.

Note that, Lemma 3.1 indicates the asymptotic equivalence of the smoothed estimating functions and their original counterpart. Theorems 3.1 and 3.2 illustrate the asymptotic equivalence of the two corresponding estimators. From Theorem 3.2, we can obtain a natural sandwich form estimator of the variance-covariance matrix of $\sqrt{m}(\tilde{\beta}_\tau - \beta_\tau)$ as

$$\widehat{\text{cov}}(\sqrt{m}(\tilde{\beta}_\tau - \beta_\tau)) = \hat{G}^{-1}(\tilde{\beta}_\tau)\tilde{V}\{\hat{G}^{-1}(\tilde{\beta}_\tau)\}^T, \tag{7}$$

where in the covariance matrix, we have $\tilde{G}(\tilde{\beta}_\tau) = \frac{1}{m}\sum_{i=1}^m X_i^T \mathit{\Gamma}_i \Sigma_i^{-1}(\rho)\mathit{\Gamma}_i X_i$ and $\tilde{V} = \frac{1}{m}\sum_{i=1}^m X_i^T \mathit{\Gamma}_i \Sigma_i^{-1}(\rho)\,\text{cov}\{\tilde{\psi}_\tau(y_i - X_i\tilde{\beta}_\tau)\}\Sigma_i^{-1}(\rho)\mathit{\Gamma}_i X_i$. Based on this formula, we update matrix $\tilde{\Omega}$ in the Newton-Raphson iteration on page 7.

Proofs are deferred to the Appendix.

4. Simulation studies

In order to examine the small sample performance of the proposed method, we conducted extensive simulation studies. A part of the simulation results are reported in this section.

The random samples are generated from the model

$$y_{ij} = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + \varepsilon_{ij} \tag{8}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, where x_{ij1} are sampled from the Bernoulli distribution with probability 0.5, $Bernoulli(0.5)$, and x_{ij2} are generated from a standard normal distribution. In this simulation study, we set the sample size $m = 500$ and a balanced design $n_i = 4$ for all $i = 1, \ldots, 500$. Let the variance-covariance matrix of ε_i with an AR(1) structure be expressed as

$$\mathit{\Sigma}_\varepsilon(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n_i-1} \\ \rho & 1 & \rho & \ldots & \rho^{n_i-2} \\ & & & \vdots & \\ \rho^{n_i-1} & \rho^{n_i-2} & & \ldots & 1 \end{pmatrix},$$

where ρ is set to be 0.1, 0.5, or 0.9 respectively, to generate errors with low, medium and high correlation. Three different distributions are considered for the random error ε_i:

Case 1. Normal distribution, assume that ε_i follows a multivariate normal distribution with mean $-q_\tau$, or the τth quantile of 0 and covariance $\mathit{\Sigma}_\varepsilon(\rho)$, $N_p(-q_\tau, \mathit{\Sigma}_\varepsilon(\rho))$, where q_τ is the τth quantile of the standard normal distribution.

Case 2. Chi-squared distribution, assume that $\varepsilon_i = \varepsilon_i' - q_\tau$, where ε_i' follows a multivariate Chi-squared distribution with two degrees of freedom (χ_2^2), where q_τ is the τth quantile of the χ_2^2 distribution.

Case 3. Student's T distribution, suppose that $\varepsilon_i = \varepsilon_i' - q_\tau$, where ε_i' follows a multivariate T distribution with three degrees of freedom (T_3), where q_τ is the τth quantile of the T_3 distribution.

The values of parameters used in the simulation are $\beta_0 = -0.5$, $\beta_1 = 0.5$ and $\beta_2 = 1$. Quantiles of $\tau = 0.25$, 0.5 and 0.95 are chosen to study the performance of the quantile regression estimators for the response distribution.

Table 1: Biases and relative efficiencies to the estimators of β_0, β_1 and β_2 using different methods at quantiles 0.25, 0.5, 0.95 are reported for Case 1 when $\rho = 0.1, 0.5, 0.9$.

			β_0		β_1		β_2	
τ	ρ	Method	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	AQR	0.0019	1.046	0.0015	1.041	-0.0001	1.069
		PQR	-0.0005	1.040	0.0014	1.044	-0.0001	1.069
		WI	0.0000	1.000	0.0020	1.000	-0.0002	1.000
	0.5	AQR	0.0043	1.098	-0.0017	1.191	0.0015	1.236
		PQR	0.0017	1.111	-0.0017	1.194	0.0015	1.242
		WI	0.0025	1.000	-0.0019	1.000	0.0016	1.000
	0.9	AQR	0.0048	1.186	-0.0006	2.811	0.0000	2.707
		PQR	0.0013	1.195	-0.0008	2.816	0.0000	2.706
		WI	0.0028	1.000	-0.0023	1.000	0.0006	1.000
0.5	0.1	AQR	0.0022	1.050	-0.0020	1.054	0.0006	1.049
		PQR	0.0022	1.050	-0.0020	1.053	0.0006	1.049
		WI	0.0019	1.000	-0.0017	1.000	0.0008	1.000
	0.5	AQR	0.0002	1.059	0.0018	1.260	0.0010	1.247
		PQR	0.0002	1.059	0.0018	1.260	0.0010	1.247
		WI	-0.0000	1.000	0.0025	1.000	0.0006	1.000
	0.9	AQR	-0.0003	1.256	-0.0005	3.135	0.0001	3.026
		PQR	-0.0003	1.256	-0.0005	3.136	0.0001	3.026
		WI	-0.0004	1.000	-0.0013	1.000	0.0015	1.000
0.95	0.1	AQR	-0.0094	1.066	0.0047	1.069	0.0029	1.099
		PQR	0.0033	1.092	0.0046	1.071	0.0028	1.118
		WI	-0.0004	1.000	0.0032	1.000	0.0031	1.000
	0.5	AQR	-0.0142	0.988	-0.0008	1.136	-0.0018	1.257
		PQR	0.0001	1.092	-0.0016	1.144	-0.0015	1.248
		WI	-0.0039	1.000	-0.0039	1.000	-0.0017	1.000
	0.9	AQR	-0.0136	1.212	-0.0003	2.152	-0.0015	2.187
		PQR	0.0037	1.244	-0.0001	2.155	-0.0017	2.129
		WI	-0.0021	1.000	0.0052	1.000	-0.0002	1.000

The results of $1,000$ simulation runs of quantile regression using different estimation methods are analyzed. We report the average bias ($Bias$) and relative efficiency (EFF) of the estimates of β_0, β_1 and β_2 using different quantile regression methods (quantile regression method assuming working independence (WI), proposed quantile regression method (PQR), and adjusted quantile regression method (AQR)) in the attached Tables. For each estimator, we use SD to denote the standard deviation of 1000 parameter estimates, SE the average of 1000 estimated standard errors. For our proposed estimators, $P_{0.95}$ denotes the percentage of simulation runs when the true parameter falls into the 95% confidence intervals constructed

Table 2: Simulation Results with Normal Errors (*case* 1).

			β_0			β_1			β_2		
τ	ρ	Method	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	AQR	0.043	0.042	0.951	0.060	0.060	0.949	0.029	0.030	0.950
		PQR	0.043	0.042	0.953	0.060	0.059	0.951	0.029	0.030	0.949
	0.5	AQR	0.047	0.046	0.947	0.055	0.055	0.947	0.028	0.027	0.952
		PQR	0.047	0.046	0.950	0.054	0.055	0.946	0.027	0.028	0.952
	0.9	AQR	0.054	0.053	0.944	0.038	0.036	0.939	0.019	0.018	0.949
		PQR	0.054	0.053	0.945	0.038	0.036	0.939	0.019	0.018	0.947
0.5	0.1	AQR	0.040	0.040	0.943	0.055	0.055	0.945	0.027	0.027	0.956
		PQR	0.040	0.040	0.943	0.055	0.055	0.945	0.027	0.027	0.956
	0.5	AQR	0.044	0.043	0.948	0.050	0.049	0.949	0.025	0.024	0.955
		PQR	0.044	0.043	0.948	0.050	0.049	0.949	0.025	0.024	0.955
	0.9	AQR	0.045	0.050	0.947	0.032	0.033	0.948	0.017	0.016	0.950
		PQR	0.045	0.050	0.947	0.032	0.033	0.948	0.017	0.016	0.950
0.95	0.1	AQR	0.066	0.062	0.926	0.090	0.088	0.945	0.046	0.043	0.934
		PQR	0.066	0.063	0.933	0.090	0.089	0.944	0.045	0.044	0.939
	0.5	AQR	0.070	0.065	0.924	0.090	0.084	0.923	0.044	0.041	0.935
		PQR	0.068	0.066	0.939	0.089	0.085	0.931	0.044	0.042	0.943
	0.9	AQR	0.080	0.079	0.926	0.063	0.058	0.919	0.031	0.028	0.928
		PQR	0.080	0.079	0.920	0.063	0.059	0.925	0.031	0.028	0.929

based on the sandwich estimate of the covariance matrix of $\hat{\beta}_\tau$, at quantiles $0.25, 0.5, 0.95$. Where ρ is specified as $0.1, 0.5$, and 0.9 respectively.

Table 1 shows the results when ε_i follows a multivariate normal distribution (*Case* 1) with an AR(1) correlation structure where the value of ρ is specified as 0.1, 0.5 and 0.9 respectively. As we can see, when the correlation is low ($\rho = 0.1$), the average biases and relative efficiencies of quantile regression estimators $\hat{\beta}_{PQR\tau}$ and $\hat{\beta}_{AQR\tau}$ are comparable, and these two estimators perform slightly better than the quantile regression estimator assuming working independence ($\hat{\beta}_{WI\tau}$). When the correlation is higher ($\rho = 0.5$, or $\rho = 0.9$), the proposed estimators $\hat{\beta}_{PQR\tau}$ and $\hat{\beta}_{AQR\tau}$ are equally efficient with small biases and much smaller variances than the working independence estimator. Moreover, the estimators $\hat{\beta}_{PQR\tau}$ and $\hat{\beta}_{AQR\tau}$ become more efficient as the correlation (ρ) increases. In general, these two proposed methods provide more efficient estimates of $\beta_{1\tau}$ and $\beta_{2\tau}$ than the intercept parameter $\beta_{0\tau}$. Similar performances are observed when ε_i is χ^2_2 (*case* 2) or T_3 (*case* 3) distributed except that the proposed estimators are more efficient at higher quantiles when the random effect follows a χ^2_2 distribution (*case* 2). The results are not reported.

Another observation was made to compare the sample standard deviation (SD) and the average asymptotic standard errors (SE) of the proposed and adjusted estimators when ε_i is normally distributed. In Table 2 we can see that each value of SD is very small and

Table 3: Simulation Results of Linear Mixed Effect Model and Median Regression Models

			β_0		β_1		β_2	
Err	ρ	Method	Bias	EFF	Bias	EFF	Bias	EFF
Nor	0.1	LME	0.0015	1.542	-0.0020	1.546	0.0004	1.543
		PQR	0.0022	1.050	-0.0020	1.053	0.0006	1.049
		WI	0.0019	1.000	-0.0017	1.000	0.0008	1.000
	0.5	LME	0.0008	1.517	0.0022	2.088	0.0006	2.048
		PQR	0.0002	1.059	0.0018	1.260	0.0010	1.247
		WI	-0.0000	1.000	0.0025	1.000	0.0006	1.000
	0.9	LME	-0.0021	1.729	0.0007	7.979	0.0001	7.605
		PQR	-0.0003	1.256	-0.0005	3.136	0.0001	3.026
		WI	-0.0004	1.000	-0.0013	1.000	0.0015	1.000
Chi	0.1	LME	0.6193	0.011	-0.0024	1.008	0.0005	1.002
		PQR	0.0084	1.030	-0.0059	1.037	0.0015	1.042
		WI	0.0069	1.000	-0.0060	1.000	0.0018	1.000
	0.5	LME	0.6159	0.011	0.0014	1.148	-0.0005	1.020
		PQR	0.0064	1.033	0.0008	1.061	-0.0004	1.084
		WI	0.0051	1.000	0.0006	1.000	0.0001	1.000
	0.9	LME	0.6149	0.018	0.0003	2.802	-0.0002	2.406
		PQR	0.0053	1.188	-0.0019	1.982	0.0004	1.864
		WI	0.0024	1.000	0.0010	1.000	0.0008	1.000
T	0.1	LME	0.0016	0.613	0.0001	0.650	-0.0007	0.631
		PQR	0.0004	1.031	-0.0007	1.052	-0.0001	1.046
		WI	0.0005	1.000	-0.0009	1.000	0.0001	1.000
	0.5	LME	0.0008	0.583	-0.0002	0.849	-0.0009	0.831
		PQR	0.0006	1.122	0.0003	1.234	0.0004	1.301
		WI	0.0005	1.000	0.0009	1.000	0.0001	1.000
	0.9	LME	0.0034	0.591	-0.0012	2.863	0.0006	3.338
		PQR	0.0020	1.223	-0.0003	2.916	0.0002	2.740
		WI	0.0031	1.000	-0.0015	1.000	0.0010	1.000

close to the corresponding SE value, which means our estimators perform very well and the estimate of the standard deviation of $\hat{\beta}_\tau$ works very well also. Furthermore, the percentages of simulation runs ($P_{0.95}$) when the true parameters fall into the 95% confidence intervals are all very close to their nominal level, evidencing the asymptotic normality of the estimators. Hence inferences based on it are reliable. The results for Cases 2 and 3 are similar.

Simulation results comparing the linear mixed effects model (LME) and the proposed median regression models are reported in Table 3. Biases($Bias$) and relative efficiencies(EFF) to each estimator are reported for three different error distributions($case$ 1, 2 and 3).

As we have expected, quantile (Median) regression outperforms mean regression when

Figure 1: Histogram of measured labor pain for all 83 women.

the random error distribution is skewed or heavy-tailed. When the error term follows a normal distribution, the LME and the proposed quantile method have comparable bias, but the LME is more efficient than the median regression according to the average of the estimated efficiencies of the three β_τ-parameters. However, when the error follows chi-square distribution (χ_2^2) or student's t distribution (T_3), the LME performs worse than our proposed median regression method, particularly in estimating the intercept parameter $\beta_{0\tau}$. The median regression model is more robust to model mis-specification, while LME can only provide misleading results in those cases.

5. A real data example

In this section, we illustrate the proposed method for quantile regression by analyzing the labor pain data, reported by Davis (1991) and successfully analyzed by Jung (1996). The data set arose from a randomized clinical trial on the effectiveness of a medication for labor pain relief. A total of $m = 83$ women were randomly assigned to either a pain medication group (43 women) or a placebo group (40 women). The response is a self-reported measure of pain measured every 30 minutes on a 100-mm line, where 0 = "no pain" and 100 = "extreme unbearable pain". The maximum number of measures for each women was 6, but at later measurement times there are numerous values missing with a nearly monotone pattern. In Figure 1, a histogram of all the pains shows that the data is severely skewed. Therefore mean regression may not be appropriate. In Figure 2, a box-plot shows the mean and median of

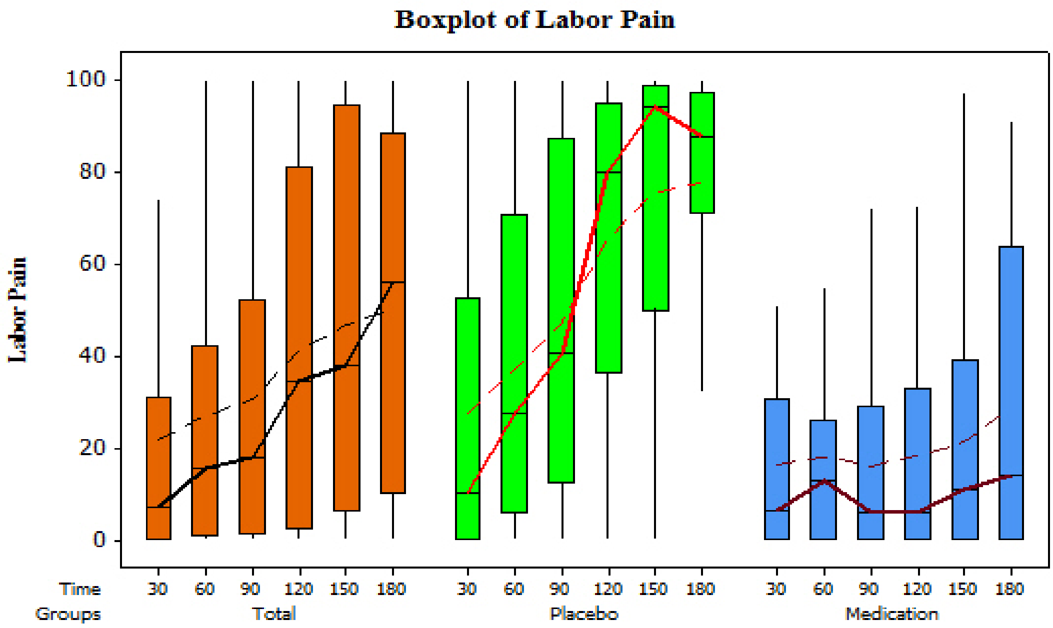

Figure 2: Box-plot of measured labor pain for all women in placebo and medication groups. The thick solid lines represent the median, while the means are connected with thin dashed lines.

the pain over time for all 83 women and those in two different groups. Statistical dependence on the temporal course of the quartiles of the response is evident to some extent, especially for the placebo group.

Let y_{ij} be the amount of pain for the ith patient at time j, R_i be the treatment indicator taking 0 for placebo and 1 for medication, and T_{ij} be the measurement time divided by 30 minutes. Jung (1996) considered the median regression model

$$y_{ij} = \beta_0 + \beta_1 R_i + \beta_2 T_{ij} + \beta_3 R_i T_{ij} + \varepsilon_{ij}, \tag{9}$$

where ε_{ij} is a zero-median error term. Note that $(\beta_0 + \beta_1) + (\beta_2 + \beta_3)T_{ij}$ is the median for the treatment group and the median for the placebo group is $\beta_0 + \beta_2 T_{ij}$.

Our proposed quantile regression model was fit for three quartiles, $\tau = 0.25, 0.5$ and 0.75, respectively. We report the estimated parameters $(\widehat{\beta})$, their asymptotic standard errors (SE) and the 95% confidence intervals (CI) in Table 4. Here we also list the results of the usual quantile regression method assuming working independence for comparison. At the 0.25th quantile, we see that our proposed method gives smaller standard errors, although these two methods produce comparable estimates of parameters. Note that all parameter estimates are significant at 5% level, meaning that each covariate has effect on the 25% quantile labor pain. Parameter estimates to the median regression methods have similar properties, except that the usual quantile regression method assuming working independence gives insignificant

Table 4: Estimated parameters ($\widehat{\beta}$), their standard errors (SE) and corresponding 95% confidence intervals (CI) from fitting both the proposed quantile regression model and usual quantile regression assuming working independence at three quartiles, $\tau = 0.25, 0.5, 0.75$.

		Proposed Method			WI		
τ	β	$\widehat{\beta}$	SE	CI	$\widehat{\beta}$	SE	CI
0.25	β_0	-10.32	0.42	(-11.13, -9.50)	-10.83	2.20	(-15.14, -6.52)
	β_1	9.08	0.42	(8.27, 9.90)	10.83	2.20	(6.51, 15.15)
	β_2	17.72	0.41	(16.92, 18.51)	10.83	2.20	(6.52, 15.14)
	β_3	-15.58	0.41	(-16.38, -14.79)	-10.83	2.20	(-15.15, -6.51)
0.5	β_0	-10.44	1.54	(-13.45, -7.43)	-6.20	7.95	(-21.77, 9.37)
	β_1	8.96	1.54	(5.95, 11.97)	12.20	8.88	(-5.21, 29.61)
	β_2	21.05	1.27	(18.56, 23.53)	17.20	2.35	(12.60, 21.80)
	β_3	-12.25	1.27	(-14.74, -9.77)	-16.20	2.72	(-21.53, -10.87)
0.75	β_0	1.02	4.08	(-6.97, 9.02)	58.67	14.83	(29.60, 87.74)
	β_1	20.42	4.08	(12.43, 28.42)	-42.67	16.30	(-74.61, -10.72)
	β_2	22.84	0.68	(21.51, 24.17)	7.67	3.44	(0.93, 14.40)
	β_3	-10.46	0.68	(-11.79, -9.13)	-2.67	4.02	(-10.54, 5.21)

estimates of β_0 and β_1, indicating similar baseline pain among two groups. While, for the third quartile (0.75th quantile), our proposed method and the WI method have very different parameter estimates with the proposed method giving much smaller standard errors of the estimates. The insignificant β_3 in WI method indicates similar time effects on the amount of pain in groups of placebo and medication, which contradicts our medical knowledge, while the significance of β_3 in our proposed method provides a perfect interpretation.

To investigate how treatment and time affect the amount of labor pain at three quartiles $(0.25, 0.5, 0.75)$, we use our proposed method to compare the estimated values of β_0 with $\beta_0 + \beta_1$ and β_2 with $\beta_2 + \beta_3$ at each quartile, respectively. The result is visualized in Figure 3. In Figure 3, we can easily see that medication treatment do help women relieve their labor pain, and the pain of women in the placebo group grows faster with time than that in the treatment group. Moreover, the amount of pain tends to grow slightly faster at higher quantiles than that at lower quantiles. These conclusions are consistent with the box plots shown in Figure 2 and results in Jung (1996) and Leng and Zhang (2012).

6. Conclusion

In this paper, we have proposed a new quantile regression model for longitudinal data, incorporating the correlations between repeated measures. We applied a general stationary auto-correlation structure to the estimating equations. To reduce the computational burden caused by the non-continuous estimating functions, we have employed the induced smoothing method of Fu and Wang (2012) for quantile regression. The estimates of the regression

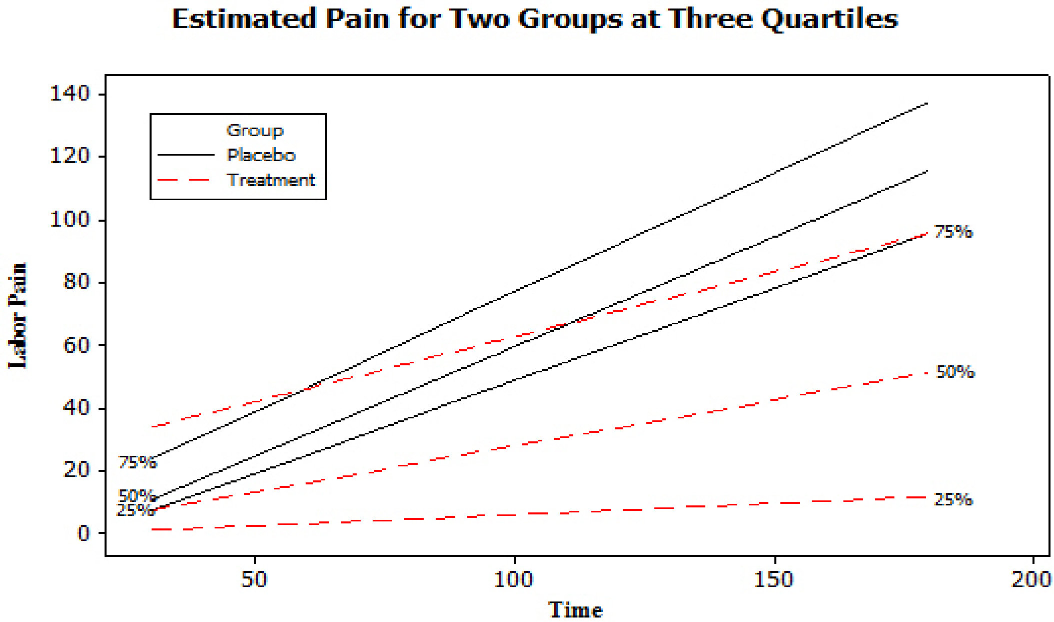

Figure 3: Labor pain obtained by using proposed quantile regression method at three quartiles 25%, 50% and 75%.

parameters and their covariance matrix are then obtained using Newton-Raphson iteration technique. It can be seen that our proposed method is a simple and efficient way to account for within-subject correlations in quantile regression for longitudinal data. This approach drew the inferential methods of quantile regression and the classical mean regression much closer. It reveals that the techniques in GEE's are applicable in quantile regression modeling. Our simulation studies indicate that the proposed method performs better than other methods assuming working independence especially when the within correlation is high. Furthermore, a comparison is also made between the proposed median regression estimator and the corresponding mean regression estimator, where the former is found to be better in analyzing heavy-tailed or skewed data. Finally, the proposed quantile regression estimator is applied to a real data set where the labor pain of two groups of women are reported, which reveals how treatment and time affect the amount of labor pain at three quartiles.

We were trying to take the within-subject correlation into consideration of quantile regression modeling, while the effects of unobserved covariates which may be different from individual to individual have not been captured. For instance, in our real data application, the personal perception of labor pain may vary from one to another. Therefore, like what has been done in Koenker (2004), we may extend our proposed model to a penalized version allowing individual specific effects by adding subject specific parameters and a penalty term. Further developments of our proposed method include extending quantile regression to well studied research areas in mean regression for longitudinal data such as mixed models for count and binary data (Sutradhar, 2011), nonlinear models (He et al., 2003), semi-parametric models (Lin and Carroll, 2006), and nonparametric models (Wu and Zhang, 2006; Qu and Li, 2006). Further results will be reported in forthcoming papers.

Acknowledgement.

The research of the authors was partially supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

References

Brown, B.M., Wang, Y.G., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrika 92, 149–158.

Chen, L., Wei, L.J., Parzen, M.I., 2004. Quantile regression for correlated observations, in: Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data, pp. 51–70.

Davis, S., 1991. Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. Statistics in Medicine 10, 1959–1980.

Farcomeni, A., 2012. Quantile regression for longitudinal data based on latent markov subject-specific parameters. Statistics and Computing 22, 141–152.

Fu, L., Wang, Y.G., 2012. Quantile regression for longitudinal data with a working correlation model. Computational Statistics and Data Analysis 56, 2526–2538.

Geraci, M., Bottai, M., 2007. Quantile regression for longitudinal data using the asymmetric laplace distribution. Biostatistics 8, 140–154.

He, X., Fu, B., Fung, W.K., 2003. Median regression of longitudinal data. Statistics in Medicine 22, 3655–3669.

Hendricks, W., Koenker, R., 1992. Hierarchical spline models for conditional quantiles and the demand for electricity. Journal of the American Statistical Association 87, 58–68.

Jung, S.H., 1996. Quasi-likelihood for median regression models. Journal of the American Statistical Association 91, 251–257.

Jung, S.H., Ying, Z., 2003. Rank-based regression with repeated measurements data. Biometrika 90, 732–740.

Karlsson, A., 2008. Nonlinear quantile regression estimation of longitudinal data. Communications in Statistics-Simulation and Computation 37, 114–131.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.

Koenker, R., 2005. Quantile Regression. Number 38 in Econometric Society monographs, New York: Combridge University Press.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. Econometrica 46, 33–55.

Koenker, R., D'Orey, V., 1987. Computing regression quantiles. Applied Statistics 36, 383–393.

Leng, C., Zhang, W., 2012. Smoothing combined estimating equations in quantile regression for longitudinal data. Statistics and Computing 17 October 2012, 1–14.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Lin, X., Carroll, R.J., 2006. Semiparametric estimation in general repeated measures problems. Journal of the Royal Statistical Society: Series B 68, 69–88.

Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., 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. Journal of the Royal Statistical Society: Series C 46, 463–476.

Liu, Y., Bottai, M., 2009. Mixed-effects models for conditional quantiles with longitudinal data. International Journal of Biostatistics 5, Article 28.

Mu, Y., Wei, Y., 2009. A dynamic quantile regression transformation model for longitudinal data. Statistica Sinica 19, 1137–1153.
Pollard, D., 1990. Empirical processes: theory and applications, in: NSF-CBMS Regional Conference Series in Probability and Statistics, Institute of Mathematical Statistics, Hayward, California.
Qu, A., Li, R., 2006. Nonparametric modeling and inference function for longitudinal data. Biometrics 62, 379–391.
Qu, A., Lindsay, B.G., Li, B., 2000. Improving generalized estimating equations using quadratic inference functions. Biometrika 87, 823–836.
Reich, B.J., Bondell, H.D., Wang, H.J., 2010. Flexible bayesian quantile regression for independent and clusterd data. Biostatistics 11, 337–352.
Sutradhar, B.C., 2003. An overview on regression models for descrete longitudinal responses. Statistical Science 18, 377–393.
Sutradhar, B.C., 2011. Dynamic Mixed Models for Familial Longitudinal Data. Springer Series in Statistics, New York: Springer. doi:10.1007/978-1-4419-8342-8.
Sutradhar, B.C., Kovacevic, M., 2000. Analyzing ordinal longitudinal survey data: generalized estimating equations approach. Biometrika 87, 837–848.
Tang, C.Y., Leng, C., 2011. Empirical likelihood and quantile regression in longitudinal data analysis. Biometrika 98, 1001–1006.
Wang, H., 2009. Inference on quantile regression for heteroscedastic mixed models. Statistica Sinica 19, 1247–1261.
Wang, H., Fygenson, M., 2009. Inference for censored quantile regression models in longitudinal studies. Annals of Statistics 37, 756–781.
Wang, H., He, X., 2007. Detecting differential expressions in genechip microarray studies: a quantile approach. Journal of the American Statistical Association 102, 104–112.
Wang, H., Zhu, Z., Zhou, J., 2009. Quantile regression in partially linear varying coefficient models. Annals of Statistics 37, 3841–3866.
Wei, Y., He, X., 2006. Conditional growth charts (with discussion). Annals of Statistics 34, 2069–2131.
Wu, H., Zhang, J.T., 2006. Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches. Wiley Series in Probability and Statistics, Wiley-Interscience.

Appendix.

In the appendix we give a set of regularity conditions and outline the proofs of the theorems in Section 3.

A1. For each i, the number of repeated measures n_i is bounded and the dimension p of covariates x_{ij} is fixed. The cumulative distribution functions $F_{ij}(z) = P(y_{ij} - x_{ij}^T\beta_\tau \leq z|x_{ij})$ are absolutely continuous, with continuous densities f_{ij} and its first derivative being uniformly bounded away from 0 and ∞ at the point 0, $i = 1, \ldots, m$; $j = 1, \ldots, n_i$.

A2. The true value β_τ is an interior point of a bounded convex region $\mathfrak{B}$.

A3. Each x_i satisfies the following conditions

(a) For any positive definite matrix W_i, $\frac{1}{m}\sum_{i=1}^m X_i^T W_i \varGamma_i X_i$ converges to a positive definite matrix; where $\varGamma_i$ is an $n_i \times n_i$ diagonal matrix with the jth diagonal element $f_{ij}(0)$.

(b) $\sup_i \|x_i\| < +\infty$, where $\|\cdot\|$ denotes the Euclidean norm.

A4. Matrix $\varOmega$ is positive definite and $\varOmega = O(\frac{1}{m})$.

A5. The differentiation of negative $\tilde{U}(\beta_\tau)$, $-\partial\tilde{U}(\beta_\tau)/\partial\beta_\tau$, is positive definite with probability 1.

Proof of Theorem 3.1. Let $H_i^T = X_i^T \mathit{\Gamma}_i \Sigma_i^{-1}(\rho)$ and $\psi_i = \psi_\tau(y_i - X_i\hat{\beta}_\tau)$, therefore $U(\hat{\beta}_\tau) = \sum_{i=1}^m H_i^T \psi_i$. Let $\bar{U}(\hat{\beta}_\tau) = \sum_{i=1}^m H_i^T \varphi_i$, where $\varphi_i = (\tau - P(y_{i1} - x_{i1}^T\hat{\beta}_\tau \le 0), \ldots, \tau - P(y_{in_i} - x_{in_i}^T\hat{\beta}_\tau \le 0))^T$. We can obtain

$$\begin{aligned}
\frac{1}{m}(U(\hat{\beta}_\tau) - \bar{U}(\hat{\beta}_\tau)) &= \frac{1}{m}\sum_{i=1}^m H_i^T(\psi_i - \varphi_i) \\
&= \frac{1}{m}\sum_{i=1}^m H_i^T \begin{pmatrix} P(y_{i1} - x_{i1}^T\hat{\beta}_\tau \le 0) - I(y_{i1} - x_{i1}^T\hat{\beta}_\tau \le 0) \\ \vdots \\ P(y_{in_i} - x_{in_i}^T\hat{\beta}_\tau \le 0) - I(y_{in_i} - x_{in_i}^T\hat{\beta}_\tau \le 0) \end{pmatrix} \\
&= \frac{1}{m}\sum_{i=1}^m \sum_{j=1}^{n_i} h_{ij}[P(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) - I(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0)],
\end{aligned}$$

where h_{ij} is a $p \times 1$ vector and $(h_{i1}, \ldots, h_{in_i}) = H_i^T$. According to the uniform strong law of large numbers (Pollard, 1990), under condition A3 we have

$$\sup_{\hat{\beta}_\tau \in \mathfrak{B}} \left| \frac{1}{m}\sum_{i=1}^m \sum_{j=1}^{n_i} h_{ij}[P(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) - I(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0)] \right| = o(m^{-1/2}) \quad \text{a.s.}.$$

Therefore,

$$\sup_{\hat{\beta}_\tau \in \mathfrak{B}} \| \frac{1}{m}(U(\hat{\beta}_\tau) - \bar{U}(\hat{\beta}_\tau))\| = o(m^{-1/2}) \quad \text{a.s.}.$$

Now,

$$G_m(\beta_\tau) = -\frac{1}{m}\frac{\partial \bar{U}(\hat{\beta}_\tau)}{\partial \hat{\beta}_\tau}\bigg|_{\hat{\beta}_\tau = \beta_\tau} = \frac{1}{m}\sum_{i=1}^m H_i^T \mathit{\Gamma}_i X_i$$

is positive definite and, with probability 1, $G_m(\beta_\tau) \to G(\beta_\tau)$ when $m \to +\infty$. Because $P(y_{ij} - x_{ij}^T\beta_\tau \le 0) = \tau$, β_τ is the unique solution of the equation $\bar{U}(\hat{\beta}_\tau)$. Together with $U(\hat{\beta}_\tau) = 0$ and condition A3, implies that $\hat{\beta}_\tau \to \beta_\tau$ as $m \to \infty$.

Because ψ_i are independent random variables with mean zero, and $\mathrm{var}\{U(\beta_\tau)/m\} = \frac{1}{m}\sum_{i=1}^m X_i^T \mathit{\Gamma}_i \Sigma_i^{-1}(\rho)\,\mathrm{cov}(\psi_i)\Sigma_i^{-1}(\rho)\mathit{\Gamma}_i X_i$, the multivariate central limit theorem implies that $\frac{1}{\sqrt{m}}U(\beta_\tau) \to N(0, V)$.

For any $\hat{\beta}_\tau$ satisfying $\|\hat{\beta}_\tau - \beta_\tau\| < cm^{-1/3}$,

$$\begin{aligned}
U(\hat{\beta}_\tau) - U(\beta_\tau) &= \sum_{i=1}^m H_i^T(\hat{\beta}_\tau)\psi_i(\hat{\beta}_\tau) - \sum_{i=1}^m H_i^T(\beta_\tau)\psi_i(\beta_\tau) \\
&= \sum_{i=1}^m H_i^T(\hat{\beta}_\tau)\{\psi_i(\hat{\beta}_\tau) - \psi_i(\beta_\tau)\} + \sum_{i=1}^m \{H_i^T(\hat{\beta}_\tau) - H_i^T(\beta_\tau)\}^T \psi_i(\beta_\tau).
\end{aligned}$$

The first term can be written as

$$\begin{aligned}
&\sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\{\psi_i(\hat{\beta}_\tau) - \psi_i(\beta_\tau)\} \\
&\quad = \sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\varphi_i(\hat{\beta}_\tau) + \sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\{\psi_i(\hat{\beta}_\tau) - \psi_i(\beta_\tau) - \varphi_i(\hat{\beta}_\tau)\} \\
&\quad = \sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\varphi_i(\hat{\beta}_\tau) + \sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\{P(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) - I(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) \\
&\qquad + I(y_{ij} - x_{ij}^T\beta_\tau \le 0) - \tau\}
\end{aligned}$$

The Lemma in Jung (1996) tells us that

$$\begin{aligned}
&\sup\left|\sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\{P(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) - I(y_{ij} - x_{ij}^T\hat{\beta}_\tau \le 0) + I(y_{ij} - x_{ij}^T\beta_\tau \le 0) - \tau\}\right| \\
&= o_p(\sqrt{m}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\{\psi_i(\hat{\beta}_\tau) - \psi_i(\beta_\tau)\} &= \sum_{i=1}^{m} H_i^T(\hat{\beta}_\tau)\varphi_i(\hat{\beta}_\tau) + o_p(\sqrt{m}) \\
&= \bar{U}(\hat{\beta}_\tau) + o_p(\sqrt{m})
\end{aligned}$$

From the law of large numbers (Pollard, 1990) the second term

$$\begin{aligned}
\sum_{i=1}^{m}\{H_i^T(\hat{\beta}_\tau) - H_i^T(\beta_\tau)\}^T\psi_i(\beta_\tau) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(h_{ij}(\hat{\beta}_\tau) - h_{ij}(\beta_\tau))[P(y_{ij} - x_{ij}^T\beta_\tau \le 0) \\
&\quad - I(y_{ij} - x_{ij}^T\beta_\tau \le 0)] \\
&= o_p(\sqrt{m}).
\end{aligned}$$

Hence, $U(\hat{\beta}_\tau) - U(\beta_\tau) = \bar{U}(\hat{\beta}_\tau) + o_p(\sqrt{m})$. Using Taylor's expansion of $\bar{U}(\hat{\beta}_\tau)$, we have

$$\frac{1}{\sqrt{m}}\{U(\hat{\beta}_\tau) - U(\beta_\tau)\} = \frac{1}{m}\frac{\partial \bar{U}(\hat{\beta}_\tau)}{\partial \hat{\beta}_\tau}\bigg|_{\hat{\beta}_\tau = \beta_\tau}\sqrt{m}(\hat{\beta}_\tau - \beta_\tau) + o_p(1).$$

Because $\hat{\beta}_\tau$ is in the $m^{-1/3}$ neighborhood of β_τ and $U(\hat{\beta}_\tau) = 0$, we have

$$\sqrt{m}(\hat{\beta}_\tau - \beta_\tau) = G_m^{-1}(\beta_\tau)\frac{1}{\sqrt{m}}U(\beta_\tau) + o_p(1).$$

Therefore $\sqrt{m}(\hat{\beta}_\tau - \beta_\tau) \to N(0, G^{-1}(\beta_\tau)V\{G^{-1}(\beta_\tau)\}^T)$ as $m \to +\infty$. $\square$

Proof of Lemma 3.1. Let $\psi_{ij}=\psi_\tau(y_{ij}-x_{ij}^T\beta_\tau)$, $\tilde{\psi}_{ij}=\tilde{\psi}_\tau(y_{ij}-x_{ij}^T\beta_\tau)$ and $d_{ij}=\varepsilon_{ij}/r_{ij}$, where $\varepsilon_{ij}=y_{ij}-x_{ij}^T\beta_\tau$, $r_{ij}=\sqrt{x_{ij}^T\Omega x_{ij}}$. Since $\tilde{\psi}_{ij}-\psi_{ij}=\operatorname{sgn}(-d_{ij})\Phi(-|d_{ij}|)$, where $\operatorname{sgn}(\cdot)$ is the sign function, we have

$$\begin{aligned}\frac{1}{\sqrt{m}}\{\tilde{U}(\beta_\tau)-U(\beta_\tau)\}&=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}X_i^T\Gamma_i\Sigma_i^{-1}(\rho)\begin{pmatrix}\operatorname{sgn}(-d_{i1})\Phi(-|d_{i1}|)\\ \vdots\\ \operatorname{sgn}(-d_{in_i})\Phi(-|d_{in_i}|)\end{pmatrix}\\&=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}\operatorname{sgn}(-d_{ij})\Phi(-|d_{ij}|),\end{aligned}$$

where z_{ij} is the jth column of $X_i^T\Gamma_i\Sigma_i^{-1}(\rho)$. Because

$$\begin{aligned}E(\tilde{\psi}_{ij}-\psi_{ij})&=\int_{-\infty}^{+\infty}\operatorname{sgn}(-d_{ij})\Phi(-|d_{ij}|)f_{ij}(\varepsilon)d\varepsilon\\&=\int_{-\infty}^{+\infty}\Phi(-|\varepsilon|/r_{ij})\{2I(\varepsilon\le 0)-1\}f_{ij}(\varepsilon)d\varepsilon\\&=r_{ij}\int_{-\infty}^{+\infty}\Phi(-|t|)\{2I(t\le 0)-1\}[f_{ij}(0)+f_{ij}^{'}(\zeta(t))r_{ij}t]dt,\end{aligned}$$

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

$$\begin{aligned}|E(\tilde{\psi}_{ij}-\psi_{ij})|&\le r_{ij}^2\int_{-\infty}^{+\infty}|t|\Phi(-|t|)|f_{ij}^{'}(\zeta(t))|dt\\&\le Mr_{ij}^2/2.\end{aligned}$$

Under regularity conditions A3 and A4, when $m\to+\infty$,

$$\left\|\frac{1}{\sqrt{m}}E\{\tilde{U}(\beta_\tau)-U(\beta_\tau)\}\right\|\le\frac{1}{\sqrt{m}}\sup_{i,j}|z_{ij}|\sum_{i=1}^{m}Mr_{ij}^2/2=o(1).$$

Moreover,

$$\frac{1}{m}\operatorname{var}\{\tilde{U}(\beta_\tau)-U(\beta_\tau)\}=\frac{1}{m}\sum_{i=1}^{m}\operatorname{var}\left\{\sum_{j=1}^{n_i}z_{ij}\operatorname{sgn}(-d_{ij})\Phi(-|d_{ij}|)\right\}.$$

By Cauchy-Schwartz inequality,

$$\begin{aligned}\frac{1}{m}\operatorname{var}\{\tilde{U}(\beta_\tau)-U(\beta_\tau)\}\le&\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}z_{ij}^T\operatorname{var}(\tilde{\psi}_{ij}-\psi_{ij})\\&+\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k\ne j}^{n_i}z_{ij}z_{ik}^T\sqrt{\operatorname{var}(\tilde{\psi}_{ij}-\psi_{ij})\operatorname{var}(\tilde{\psi}_{ik}-\psi_{ik})}.\end{aligned}$$

Hence for each $j = 1, \dots, n_i$,

$$\begin{aligned}
\operatorname{var}(\tilde{\psi}_{ij} - \psi_{ij}) \le E(\tilde{\psi}_{ij} - \psi_{ij})^2 &= \int_{-\infty}^{+\infty} \{\operatorname{sgn}(-d_{ij})\Phi(-|d_{ij}|)\}^2 f_{ij}(\varepsilon) d\varepsilon \\
&= r_{ij} \int_{-\infty}^{+\infty} \Phi^2(-|t|) f_{ij}(r_{ij} t) dt \\
&= r_{ij} \int_{|t|>\Delta} \Phi^2(-|t|) f_{ij}(r_{ij} t) dt + r_{ij} \int_{|t|\le\Delta} \Phi^2(-|t|) f_{ij}(r_{ij} t) dt \\
&\le \Phi^2(-\Delta) + r_{ij}\Delta f_{ij}(\zeta),
\end{aligned}$$

where Δ is a positive value, and ζ is in the interval $(-r_{ij}\Delta, r_{ij}\Delta)$. Let $\Delta = m^{1/3}$. Under condition A4, because $r_{ij} = O(m^{-1/2})$, we have $r_{ij}\Delta = O(m^{-1/6})$. Moreover, both $\Phi^2(-\Delta)$ and $r_{ij}\Delta f_{ij}(\zeta)$ converges to 0 as $m \to +\infty$. By conditions A2 and A3, it can be easily obtained that $\frac{1}{m} \operatorname{var}\{\tilde{U}(\beta_\tau) - U(\beta_\tau)\} = o(1)$. Therefore, for any β_τ, we have $\frac{1}{\sqrt{m}}\{\tilde{U}(\beta_\tau) - U(\beta_\tau)\} \to 0$ as $m \to +\infty$. $\square$

Proof of Theorem 3.2. From the results in Theorem 3.1 along with $\sup_{\hat{\beta}_\tau \in \mathfrak{B}} \| m^{-1}\{U(\hat{\beta}_\tau) - \bar{U}(\hat{\beta}_\tau)\} \| = o(m^{-1/2})$ a.s., and by the triangle inequality, we have $\sup_{\hat{\beta}_\tau \in \mathfrak{B}} \| m^{-1}\{\tilde{U}(\hat{\beta}_\tau) - \bar{U}(\hat{\beta}_\tau)\} \| = o(m^{-1/2})$. If we denote β_τ as the unique solution of equation $\bar{U}(\hat{\beta}_\tau) = 0$ and $\tilde{\beta}_\tau$ solving $\tilde{U}(\hat{\beta}_\tau) = 0$, we can obtain that $\tilde{\beta}_\tau \to \beta_\tau$ as $m \to +\infty$.

Before proving the asymptotic normality of $\tilde{\beta}_\tau$, we first prove that $m^{-1}\{\tilde{G}(\beta_\tau) - G(\beta_\tau)\} \xrightarrow{p} 0$, where $\tilde{G}(\beta_\tau) = -\partial \tilde{U}(\beta_\tau)/\partial \beta_\tau = \sum_{i=1}^m X_i^T \Gamma_i \Sigma_i^{-1}(\rho) \tilde{\Lambda}_i X_i$. If we denote $H_i^T = X_i^T \Gamma_i \Sigma_i^{-1}(\rho) = (h_{i1}, \dots, h_{in_i})$, where h_{ij} is a $p \times 1$ vector, we can obtain that

$$E\{\tilde{G}(\beta_\tau)\} - G(\beta_\tau) = \sum_{i=1}^m \sum_{j=1}^{n_i} h_{ij} \left\{ \frac{1}{r_{ij}} E\phi\left(\frac{\varepsilon_{ij}}{r_{ij}}\right) - f_{ij}(0) \right\} x_{ij}.$$

Because

$$\begin{aligned}
\left| \frac{1}{r_{ij}} E\phi\left(\frac{\varepsilon_{ij}}{r_{ij}}\right) - f_{ij}(0) \right| &= \left| \frac{1}{r_{ij}} \int_{-\infty}^{+\infty} \phi\left(\frac{\varepsilon}{r_{ij}}\right) f_{ij}(\varepsilon) d\varepsilon - f_{ij}(0) \right| \\
&= \left| \int_{-\infty}^{+\infty} \phi(t) \{f_{ij}(0) + r_{ij} t f_{ij}(\xi_t)\} dt - f_{ij}(0) \right| \\
&= \left| r_{ij} \int_{-\infty}^{+\infty} \phi(t) t f_{ij}(\xi_t) dt \right| \\
&\le r_{ij} \int_{-\infty}^{+\infty} |\phi(t) t f_{ij}(\xi_t)| dt,
\end{aligned}$$

where ξ_t lies between 0 and $r_{ij} t$. By condition A1, there exists a constant M such that $f_{ij}(\xi_t) \le M$. Furthermore, according to condition A4, we have

$$\left| \frac{1}{r_{ij}} E\phi\left(\frac{\varepsilon_{ij}}{r_{ij}}\right) - f_{ij}(0) \right| \le \sqrt{\frac{2}{\pi}} r_{ij} M \to 0.$$

By the strong law of large numbers, we know that $m^{-1}\tilde{G}(\beta_\tau) \rightarrow E\{m^{-1}\tilde{G}(\beta_\tau)\}$. Using the triangle inequality, we have

$$|m^{-1}\{\tilde{G}(\beta_\tau) - G(\beta_\tau)\}| \leq |m^{-1}\{\tilde{G}(\beta_\tau) - E\tilde{G}(\beta_\tau)\}| + |m^{-1}\{E\tilde{G}(\beta_\tau) - G(\beta_\tau)\}| \rightarrow o(1),$$

which is equivalent to $m^{-1}\{\tilde{G}(\beta_\tau) - G(\beta_\tau)\} \xrightarrow{p} 0$.

By Taylor series expansion of $\tilde{U}(\hat{\beta}_\tau)$ around β_τ gives us

$$\tilde{U}(\hat{\beta}_\tau) = \tilde{U}(\beta_\tau) - \tilde{G}(\hat{\beta}_\tau^*)(\hat{\beta}_\tau - \beta_\tau),$$

where $\hat{\beta}_\tau^*$ lies between $\hat{\beta}_\tau$ and β_τ. Let $\hat{\beta}_\tau = \tilde{\beta}_\tau$. Because $\tilde{U}(\tilde{\beta}_\tau) = 0$ and $\tilde{\beta}_\tau \rightarrow \beta_\tau$, we therefore obtain $\hat{\beta}_\tau^* \rightarrow \beta_\tau$ and $\tilde{G}(\hat{\beta}_\tau^*) \rightarrow \tilde{G}(\beta_\tau)$. By Lemma 3.1 and $m^{-1}\{\tilde{G}(\beta_\tau) - G(\beta_\tau)\} \xrightarrow{p} 0$, we thus have

$$\sqrt{m}(\tilde{\beta}_\tau - \beta_\tau) = G_m^{-1}(\beta_\tau)\frac{1}{\sqrt{m}}U(\beta_\tau) + o_p(1).$$

Therefore $\sqrt{m}(\tilde{\beta}_\tau - \beta_\tau) \rightarrow N(0, G^{-1}(\beta_\tau)V\{G^{-1}(\beta_\tau)\}^T)$ as $m \rightarrow +\infty$. $\square$