SCIENCE CHINA
Mathematics

• ARTICLES •

January 2017 Vol. 60 No. 1: 147–164
doi: 10.1007/s11425-015-0175-y

Weighted quantile regression for longitudinal data using empirical likelihood

YUAN XiaoHui[1], LIN Nan[2], DONG XiaoGang[1] & LIU TianQing[3,*]

[1]*School of Basic Science, Changchun University of Technology, Changchun* 130012*, China;*
[2]*Department of Mathematics, Washington University in St. Louis, St. Louis, WA* 63130*, USA;*
[3]*School of Mathematics, Jilin University, Changchun* 130012*, China*
Email: xiaohuiyuan@outlook.com, nlin@wustl.edu, dongxiaogang@ccut.edu.cn, tianqingliu@outlook.com

Received March 14, 2015; accepted April 5, 2016; published online August 31, 2016

Abstract This paper proposes a new weighted quantile regression model for longitudinal data with weights chosen by empirical likelihood (EL). This approach efficiently incorporates the information from the conditional quantile restrictions to account for within-subject correlations. The resulted estimate is computationally simple and has good performance under modest or high within-subject correlation. The efficiency gain is quantified theoretically and illustrated via simulation and a real data application.

Keywords empirical likelihood, estimating equation, influence function, longitudinal data, weighted quantile regression

MSC(2010) 62F11, 62H11, 62J08

Citation: Yuan X H, Lin N, Dong X G, et al. Weighted quantile regression for longitudinal data using empirical likelihood. Sci China Math, 2017, 60: 147–164, doi: 10.1007/s11425-015-0175-y

1 Introduction

Since the seminal work of Koenker and Bassett [15], quantile regression [14] has emerged as an important alternative to mean regression, for that it does not require specification of the error distribution and provides a more complete description of the conditional distribution of the response variable given the covariates. See [2, 6, 8, 12, 14, 18, 23] and references therein for overview and discussions.

Longitudinal data are very common in biological and medical research. The characteristics of a longitudinal study is that repeated observations for an individual subject tend to be correlated. In the context of mean regression with longitudinal data, incorporating the within subject correlations is essential for efficient inferences, as demonstrated by Liang and Zeger [16] and Qu et al. [22], among others. However, incorporating the within subject correlations is more difficult for quantile regression in longitudinal data analysis.

Jung [9] first considered the quantile regression model for longitudinal data and proposed a quasi-likelihood method for median regression. He et al. [7] developed a generalized estimating equations approach for balanced longitudinal data, where the covariances involved in the generalized estimating equations are estimated by sample versions. Their method works well when the number of repeated measurements is small but may lead to a loss of efficiency for parameter estimation when the number of repeated measurements is large. Chen et al. [2] and Yin and Cai [28] proposed using the generalized

*Corresponding author

estimating equations with the independence working model to estimate quantile regression models, which are simple but could lead to a loss of efficiency for parameter estimation when strong correlations exist. Koenker [13] proposed using a penalized likelihood to make inferences for a quantile regression model with subject specified fixed effects. Geraci and Bottai [6] studied a quantile regression model with random intercepts to account for the within subject correlations and introduced a likelihood-based approach by assuming the response variable following an asymmetric Laplace density. Liu and Bottai [17] extended the model by Geraci and Bottai [6] to linear mixed-effects models, where random regression coefficients are specified by a multivariate Laplace distribution. Farcomeni [4] considered a quantile regression model with time-varying random effects and the random effects are modeled by a first order latent Markov chain. Wang and Zhu [27] proposed a smoothed empirical likelihood (EL; see [20]) inference procedures under the framework of quantile regression with longitudinal data, which avoids estimating the unknown error density function and the within-subject correlations involved in the asymptotic covariance matrix of quantile estimators. Fu and Wang [5] generalized the induced smoothing method (see [1]) to the quantile regression with longitudinal data, where the induced smoothing method is introduced to obtain parameter estimates and their variances from non-smooth estimating functions.

In this paper, we consider quantile regression for longitudinal data and propose a weighted quantile regression method to incorporate within-subject correlation with weights chosen by empirical likelihood (EL). Our idea was motivated by a recent study by Tang and Leng [26], who considered a special case where an auxiliary conditional mean model is available in addition to the quantile model. In practice, the method of Tang and Leng [26] will be limited when the mean regression model cannot be reliably established or the error distribution is heavy-tail with no finite mean. However, their usage of EL weights to effectively account for auxiliary information sheds important insights on improving efficiency. In this paper, we focus on quantile regression at a given quantile and adopt the idea of EL weights to borrow information from structures assumed on the conditional quantile. We show both theoretically and empirically that this approach can serve as a generic device to improve efficiency in quantile regression for longitudinal data.

2 Conventional quantile regression

Consider a longitudinal study with N subjects. For the i-th subject, $i = 1, \ldots, N$, let y_{it} and $x_{it} = (x_{it1}, \ldots, x_{itp})^{\mathrm{T}} \in \mathbb{R}^p$ be the response and covariate at time $t = 1, \ldots, n_i$. We assume that the subjects are independent and the responses for the same subject are correlated over time. For the i-th subject, let $y_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$ and $x_i = (x_{i1}, \ldots, x_{in_i})$. Suppose that the τ-th conditional quantile of y_{it} given x_i is given by $Q_\tau(y_{it} \mid x_i) = x_{it}^{\mathrm{T}}\beta_0$, where β_0 is interior to the parameter space Θ, a compact subset of $\mathbb{R}^p$. The basic approach to estimating β_0 is the conventional quantile regression (CQ) estimator,

$$\hat{\beta} = \arg\min_{\beta\in\Theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \rho_\tau(y_{it} - x_{it}^{\mathrm{T}}\beta),$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the check loss function and $I(\cdot)$ is the indicator function. In longitudinal data, the CQ estimator is consistent but typical of low relative efficiency for completely ignoring within-subject correlations. The asymptotic normality of the CQ estimator is given by Wang and Zhu [27] and included below.

Theorem 2.1. *Let $S_{it}(\beta) = I(y_{it} - x_{it}^{\mathrm{T}}\beta \leqslant 0) - \tau$, $S_i(\beta) = (S_{i1}(\beta), \ldots, S_{in_i}(\beta))^{\mathrm{T}}$, and $f_{it}(\cdot)$ be the conditional probability density function (p.d.f.) of $\varepsilon_{it} = y_{it} - x_{it}^{\mathrm{T}}\beta_0$ given x_i. Under Conditions* C1–C6 *(see Appendix), one has $\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma)$ as $N \to \infty$, where $\Sigma = D_1^{-1} D_0 D_1^{-1}$, $D_0 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i \Lambda_i x_i^{\mathrm{T}}$, $\Lambda_i = E\{S_i(\beta_0)S_i^{\mathrm{T}}(\beta_0) \mid x_i\}$, $D_1 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i \Psi_i x_i^{\mathrm{T}}$ and $\Psi_i = \mathrm{diag}\{f_{i1}(0), \ldots, f_{in_i}(0)\}$.*

One may construct a consistent estimate of the asymptotic covariance matrix Σ using the induced smoothing method [1]. Let $\hat{\varepsilon}_{it} = y_{it} - x_{it}^{\mathrm{T}}\hat{\beta}$, $\sigma_{it}^2 = x_{it}^{\mathrm{T}} x_{it}/N$, $\hat{\Psi}_{it} = \frac{1}{\sigma_{it}}\phi(\frac{\hat{\varepsilon}_{it}}{\sigma_{it}})$, $\hat{\Psi}_i = \mathrm{diag}\{\hat{\Psi}_{i1}, \ldots, \hat{\Psi}_{in_i}\}$,

$\hat{\Lambda}_i = S_i(\hat{\beta})S_i^{\mathrm{T}}(\hat{\beta})$, $\hat{D}_0 = \frac{1}{N}\sum_{i=1}^N x_i\hat{\Lambda}_i x_i^{\mathrm{T}}$ and $\hat{D}_1 = \frac{1}{N}\sum_{i=1}^N x_i\hat{\Psi}_i x_i^{\mathrm{T}}$, where $\phi(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ is the density function of the standard normal variable. Then, $\hat{\Sigma} = \hat{D}_1^{-1}\hat{D}_0\hat{D}_1^{-1}$ is a consistent estimator of Σ.

3 Weighted quantile regression

Weighting is an effective approach to improving efficiency in marginal models [24, 25]. In this section, we propose weighted quantile regression estimators in the form of

$$\hat{\beta}_{\mathrm{WQR}} = \arg\min_{\beta\in\Theta}\sum_{i=1}^N w_i\left\{\sum_{t=1}^{n_i}\rho_\tau(y_{it} - x_{it}^{\mathrm{T}}\beta)\right\}, \tag{3.1}$$

where the weights w_i's are determined by the data and reflect within-subject correlations.

3.1 Naive EL weights

Within-subject correlation serves as important auxiliary information in longitudinal data. Because quantile regression does not require a likelihood, we use estimating equation and propose to find the optimal weights using empirical likelihood. We incorporate within-subject correlations using working correlation matrices specified via the decomposition by Qu et al. [22], i.e., $R_i^{-1} = a_0 I_{n_i} + \sum_{k=1}^m a_k M_{ik}$, where $a_0, a_1, \ldots, a_m$ are unknown constants, I_d is the $d\times d$ identity matrix, and $M_{i1}, \ldots, M_{im}$ are some known non-identity matrices. Then define

$$g(z_i,\beta) = (I_m \otimes x_i)M_i S_i(\beta), \quad i = 1,\ldots,N, \tag{3.2}$$

where $z_i = (y_i, x_i^{\mathrm{T}})$, $M_i = (M_{i1}^{\mathrm{T}}, \ldots, M_{im}^{\mathrm{T}})^{\mathrm{T}}$ is an $mn_i\times n_i$ matrix, and $\otimes$ denotes the Kronecker product. Note that, when m or p is large, it is inappropriate to base $g(z_i,\beta)$ on all matrices $M_{i1}, \ldots, M_{im}$, then we may use only a subset of $\{M_{i1}, \ldots, M_{im}\}$ in M_i.

By the quantile model assumptions and Theorem 2.1, it follows that $E\{g(z_i,\beta_0)\} = 0$ and the CQ estimator $\hat{\beta}$ is a consistent estimator of β_0. Thus, we can apply the EL method to the moment restriction $\lim_{N\to\infty} E\{g(z_i,\hat{\beta})\} = 0$. Formally, let $L_{\mathrm{EL}} = \prod_{i=1}^N p_i$ with nonnegative jump sizes p_i's that sum to 1. The naive EL weights are then obtained by maximizing $\log L_{\mathrm{EL}}$ subject to the constraints

$$\sum_{i=1}^N p_i = 1, \quad \sum_{i=1}^N p_i g(z_i,\hat{\beta}) = 0, \quad p_i \geqslant 0, \quad i = 1,\ldots,N, \tag{3.3}$$

where $\hat{\beta}$ is the CQ estimator. The problem can be solved by using Lagrange multipliers. This is equivalent to maximizing

$$Q = \sum_{i=1}^N \log(p_i) - \nu\left(\sum_{i=1}^N p_i - 1\right) - n\lambda^{\mathrm{T}}\sum_{i=1}^N p_i g(z_i,\hat{\beta}),$$

where ν is the multipliers for the second set of constraints, and λ the Lagrange multipliers for the third set of constraints. The first-order-conditions of Q with respect to p_i, ν and λ are

$$\begin{aligned}
&\frac{1}{p_i} - \nu - N\lambda^{\mathrm{T}} g(z_i,\hat{\beta}) = 0, \quad i = 1,\ldots,N,\\
&\sum_{i=1}^N p_i = 1,\\
&\sum_{i=1}^N p_i g(z_i,\hat{\beta}) = 0.
\end{aligned}$$

Multiplying the first equation by p_i, summing over i and using the second and third equations, we have $\nu = N$,

$$\hat{p}_i = \frac{1}{N(1+\hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta}))}, \quad i=1,\ldots,N, \tag{3.4}$$

and $\hat{\lambda}$ solves

$$\sum_{i=1}^{N} \frac{g(z_i,\hat{\beta})}{1+\hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta})} = 0. \tag{3.5}$$

Using $\{\hat{p}_i\}_{i=1}^N$ in (3.4) as weights in (3.1), the weighted quantile regression estimator with naive EL weights is given by

$$\hat{\beta}_{\mathrm{NEL}} = \arg\min_{\beta\in\Theta} \sum_{i=1}^{N} \hat{p}_i \bigg\{ \sum_{t=1}^{n_i} \rho_\tau(y_{it} - x_{it}^{\mathrm{T}}\beta) \bigg\}. \tag{3.6}$$

Implementing $\hat{\beta}_{\mathrm{NEL}}$ is straightforward. From (3.5), it is easily seen that

$$\hat{\lambda} = \arg\max_{\lambda\in R^{mp}} \sum_{i=1}^{N} \log\{1+\lambda^{\mathrm{T}} g(z_i,\hat{\beta})\}.$$

As the objective function is globally concave, this optimization problem can be solved by a simple Newton-Raphson numerical procedure. After having $\hat{\lambda}$ and hence $\hat{p}_i$'s, we can obtain $\hat{\beta}_{\mathrm{NEL}}$ straightforwardly using the R package `quantreg`. The following theorem gives the asymptotic distribution of $\hat{\beta}_{\mathrm{NEL}}$.

Theorem 3.1. *Under the regularity conditions* C1–C6*, as $N \to \infty$, we have*

$$\sqrt{N}(\hat{\beta}_{\mathrm{NEL}} - \beta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{NEL}}),$$

where $\Sigma_{\mathrm{NEL}} = \Sigma - D_1^{-1}\Omega_1 D_1^{-1}$,

$$\begin{aligned}
\Omega_1 &= \Gamma_1\Pi^{-1}\Gamma_1^{\mathrm{T}} - D_0D_1^{-1}D_2^{\mathrm{T}}\Pi^{-1}\Gamma_1^{\mathrm{T}} - (D_0D_1^{-1}D_2^{\mathrm{T}}\Pi^{-1}\Gamma_1^{\mathrm{T}})^{\mathrm{T}} + \Gamma_1\Pi^{-1}\Gamma_1^{\mathrm{T}}D_1^{-1}D_2^{\mathrm{T}}\Pi^{-1}\Gamma_1^{\mathrm{T}} \\
&\quad + (\Gamma_1\Pi^{-1}\Gamma_1^{\mathrm{T}}D_1^{-1}D_2^{\mathrm{T}}\Pi^{-1}\Gamma_1^{\mathrm{T}})^{\mathrm{T}} - \Gamma_1\Pi^{-1}D_2D_1^{-1}D_0D_1^{-1}D_2^{\mathrm{T}}\Pi^{-1}\Gamma_1^{\mathrm{T}}, \\
D_2 &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}(I_m\otimes x_i)M_i\Psi_i x_i^{\mathrm{T}}, \\
\Gamma_1 &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i\Lambda_i M_i^{\mathrm{T}}(I_m\otimes x_i)^{\mathrm{T}}, \\
\Pi &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\{(I_m\otimes x_i)M_i\}\Lambda_i\{(I_m\otimes x_i)M_i\}^{\mathrm{T}}
\end{aligned}$$

with Ψ_i, Λ_i, D_0, D_1 and Σ defined in Theorem 2.1.

The asymptotic covariance matrix Σ_{NEL} can be estimated using the induced smoothing method. Let $\hat{\Pi} = \frac{1}{N}\sum_{i=1}^{N}\{(I_m\otimes x_i)M_i\}\hat{\Lambda}_i\{(I_m\otimes x_i)M_i\}^{\mathrm{T}}$, $\hat{D}_2 = \frac{1}{N}\sum_{i=1}^{N}(I_m\otimes x_i)M_i\hat{\Psi}_i x_i^{\mathrm{T}}$, $\hat{\Gamma}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i\hat{\Lambda}_i M_i^{\mathrm{T}}(I_m \otimes x_i)^{\mathrm{T}}$ and

$$\begin{aligned}
\hat{\Omega}_1 &= \hat{\Gamma}_1\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}} - \hat{D}_0\hat{D}_1^{-1}\hat{D}_2^{\mathrm{T}}\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}} - (\hat{D}_0\hat{D}_1^{-1}\hat{D}_2^{\mathrm{T}}\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}})^{\mathrm{T}} + \hat{\Gamma}_1\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}}\hat{D}_1^{-1}\hat{D}_2^{\mathrm{T}}\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}} \\
&\quad + (\hat{\Gamma}_1\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}}\hat{D}_1^{-1}\hat{D}_2^{\mathrm{T}}\hat{\Pi}^{-1}\hat{\Gamma}_1^{\mathrm{T}})^{\mathrm{T}} - \hat{\Gamma}_1\hat{\Pi}^{-1}\hat{D}_2\hat{D}_1^{-1}\hat{D}_0\hat{D}_1^{-1}\hat{D}_2^{\mathrm{T}}\Pi^{-1}\hat{\Gamma}_1^{\mathrm{T}},
\end{aligned}$$

where $\hat{\Psi}_i$, $\hat{\Lambda}_i$, $\hat{D}_0$ and $\hat{D}_1$ are previously defined. A consistent estimator of Σ_{NEL} is then $\hat{\Sigma}_{\mathrm{NEL}} = \hat{\Sigma} - \hat{D}_1^{-1}\hat{\Omega}_1\hat{D}_1^{-1}$.

It is worth noting that the matrix Ω_1 may not be nonnegative definite, so the efficiency gain of $\hat{\beta}_{\mathrm{NEL}}$ over the CQ estimator may not be evident theoretically. This consideration motivates our further adjustment of the EL weights.

3.2 Adjusted EL weights

Lemma A.2 in Appendix gives the following representation:

$$N^{-1/2}\sum_{i=1}^{N} g(z_i,\hat{\beta}) = N^{-1/2}\sum_{i=1}^{N}\{(I_m\otimes x_i)M_i - D_2D_1^{-1}x_i\}S_i(\beta_0)+o_p(1),$$

where D_1 and D_2 are defined in Theorems 2.1 and 3.1, respectively. This motivates us to adjust the score function in (3.2) to

$$\tilde{g}(z_i,\beta) = \{(I_m\otimes x_i)\,M_i - \hat{D}_2\hat{D}_1^{-1}x_i\}S_i(\beta), \quad i=1,\ldots,N, \tag{3.7}$$

where $\hat{D}_1$ and $\hat{D}_2$ are previously defined. Then we obtain the adjusted EL weights $\hat{q}_i$ in the same way as $\hat{p}_i$ in (3.4), and denote by $\hat{\beta}_{\mathrm{AEL}}$ the resulting weighted quantile regression estimator with adjusted EL weights.

Theorem 3.2. *Under the regularity conditions* C1–C6, *and as* $N\to\infty$,

$$\sqrt{N}(\hat{\beta}_{\mathrm{AEL}}-\beta_0)\xrightarrow{d} N(0,\Sigma_{\mathrm{AEL}}),$$

where $\Sigma_{\mathrm{AEL}} = \Sigma - D_1^{-1}\Omega_2D_1^{-1}$, $\Omega_2=\Gamma_2\Delta^{-1}\Gamma_2^{\mathrm{T}}$, $\Gamma_2=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N x_i\Lambda_iH_i^{\mathrm{T}}$, $\Delta=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N H_i\Lambda_iH_i^{\mathrm{T}}$, $H_i=(I_m\otimes x_i)M_i-D_2D_1^{-1}x_i$, *and* Λ_i, D_1, D_2 *and* Σ *are defined in Theorems* 2.1 *and* 3.1.

Since the matrix Ω_2 is obviously nonnegative definite, $\hat{\beta}_{\mathrm{AEL}}$ has an asymptotically variance no greater than that of the CQ estimator. The asymptotic covariance matrix Σ_{AEL} can also be estimated using the induced smoothing method. Let $\hat{H}_i=(I_m\otimes x_i)M-\hat{D}_2\hat{D}_1^{-1}x_i$, $\hat{\Gamma}_2=\frac{1}{N}\sum_{i=1}^N x_i\hat{\Lambda}_i\hat{H}_i^{\mathrm{T}}$, $\hat{\Delta}=\frac{1}{N}\sum_{i=1}^N\hat{H}_i\hat{\Lambda}_i\hat{H}_i^{\mathrm{T}}$ and $\hat{\Omega}_2=\hat{\Gamma}_2\hat{\Delta}^{-1}\hat{\Gamma}_2^{\mathrm{T}}$. Then, a consistent estimator of Σ_{AEL} is $\hat{\Sigma}_{\mathrm{AEL}}=\hat{\Sigma}-\hat{D}_1^{-1}\hat{\Omega}_2\hat{D}_1^{-1}$.

4 Simulation studies

In this section, we conduct a simulation study to investigate the finite-sample performance of the proposed estimators. We simulate 1,000 Monte Carlo samples of size $N=100$ from the following model under three different setups:

$$y_{it}=\mu+x_{it}^{\mathrm{T}}\beta+\varepsilon_{it}(\tau), \quad i=1,\ldots,N, \quad t=1,\ldots,n,$$

where $x_{it}=(x_{it1},x_{it2})^{\mathrm{T}}$, $\beta=(\beta_1,\beta_2)^{\mathrm{T}}$, $\varepsilon_i(\tau)=(\varepsilon_{i1}(\tau),\ldots,\varepsilon_{in}(\tau))^{\mathrm{T}}$ with $\varepsilon_{ij}(\tau)$ obtained from different transformations of latent random errors

$$\xi_i=(\xi_{i1},\ldots,\xi_{in})^{\mathrm{T}}\sim N(0,V(\alpha)).$$

The latent correlation matrix $V(\alpha)=\{V_{jk}(\alpha)\}$ is either an AR(1) ($V_{jk}(\alpha)=\alpha^{|j-k|}$) or exchangeable ($V_{jk}(\alpha)=\alpha+(1-\alpha)\,I(j=k)$) correlation matrix.

• Setup 1. $\varepsilon_{it}(\tau)=F_\tau^{-1}(\Phi(\xi_{it}))$ with F_τ being the cdf of the asymmetric Laplace distribution ALD$(0,1,\tau)$ (see [29]).

• Setup 2. $\varepsilon_{it}(\tau)=F_c^{-1}(\Phi(\xi_{it}))-F_c^{-1}(\tau)$ with F_c as the standard Cauchy cdf, i.e., $F_c(x)=0.5+\arctan(x)/\pi$. Note that the mean and variance of y_{it} do not exist under this setup.

• Setup 3. $\varepsilon_{it}(\tau)=F_{T(2)}^{-1}(\Phi(\xi_{it}))-F_{T(2)}^{-1}(\tau)$ with $F_{T(2)}$ as the cdf of Student's t-distribution with 2 degrees of freedom, i.e., $F_{T(2)}(x)=0.5+0.5x/(2+x^2)^{1/2}$. Note that the mean of y_{it} exists, but variance does not.

Note that Setups 2 and 3 present theoretical difficulty to impose an auxiliary mean regression model, so the QREL estimator in [26] does not apply. Under all three setups, we set $\tau=0.7$, $n=10$, $(\mu,\beta_1,\beta_2)=(0,1,1)$ and generate $x_{i11},x_{i21},\ldots,x_{in1}$ independently from $N(0,1)$ and $x_{i12},x_{i22},\ldots,x_{in2}$ from Bernoulli(0.5). Three levels of within-subject correlation is considered by setting $\alpha=0.3$, 0.7 and 0.9.

Table 1 Comparison of relative efficiency (RE). QREL is based on mean model $y \sim x_1 + x_2$; QREL2 is based on mean model $y \sim x_1$; HFF is the estimator in [7, Section 2.2]; AR, autoregressive; EX, exchangeable

			$\alpha = 0.9$		$\alpha = 0.7$		$\alpha = 0.3$	
Setup	$V(\alpha)$	Method	β_1	β_2	β_1	β_2	β_1	β_2
1	EX	NEL	3.352	3.044	1.585	1.745	1.075	1.134
		AEL	3.336	3.043	1.559	1.728	1.057	1.109
		QREL	1.532	1.451	1.293	1.438	1.050	1.142
		QREL2	1.554	0.959	1.307	1.004	1.065	0.974
		HFF	2.511	2.271	1.417	1.541	0.975	1.019
	AR	NEL	1.808	1.939	1.144	1.154	0.996	0.997
		AEL	1.783	1.868	1.124	1.150	0.974	1.015
		QREL	1.404	1.476	1.126	1.120	0.999	0.987
		QREL2	1.428	0.992	1.126	0.988	1.022	0.986
		HFF	2.244	2.295	1.285	1.275	0.968	0.975
2	EX	NEL	3.214	2.939	1.818	1.717	1.121	1.141
		AEL	3.099	2.832	1.725	1.687	1.130	1.109
		QREL	1.133	1.003	1.102	1.102	0.995	1.022
		QREL2	1.092	0.994	1.070	0.980	1.004	0.948
		HFF	2.631	2.469	1.776	1.645	1.061	1.115
	AR	NEL	1.781	1.823	1.177	1.153	1.013	0.995
		AEL	1.799	1.764	1.146	1.157	0.985	0.981
		QREL	1.054	1.064	0.971	0.996	0.969	0.961
		QREL2	1.058	0.977	0.984	0.980	0.996	0.964
		HFF	2.264	2.300	1.271	1.222	0.999	1.048
3	EX	NEL	3.227	3.350	1.760	1.678	1.135	1.081
		AEL	3.127	3.251	1.757	1.641	1.122	1.065
		QREL	1.433	1.577	1.374	1.249	1.123	1.085
		QREL2	1.396	0.980	1.323	0.973	1.138	0.987
		HFF	2.371	2.489	1.551	1.492	1.066	1.020
	AR	NEL	1.563	1.755	1.099	1.207	0.971	0.966
		AEL	1.571	1.762	1.103	1.186	0.960	0.934
		QREL	1.282	1.367	1.069	1.138	0.985	0.969
		QREL2	1.290	0.972	1.089	0.974	1.010	0.989
		HFF	2.164	2.218	1.382	1.387	0.969	1.004

Five estimators of β are considered. The first one is the NEL estimator proposed in Subsection 3.1. For the NEL estimator, we use $g(z_i, \beta) = x_i M S_i(\beta)$ to obtain the naive EL weights $\hat{p}_i$, $i = 1, \ldots, N$. The second one is the AEL estimator proposed in Subsection 3.2. For the AEL estimator, we use

$$\tilde{g}(z_i, \beta) = (x_i M - \hat{D}_2 \hat{D}_1^{-1} x_i) S_i(\beta)$$

to obtain the adjusted EL weights $\hat{q}_i$, $i = 1, \ldots, N$. Following [22], we take M as either a matrix with 0 on the diagonal and 1 elsewhere, or a matrix with two main off-diagonals being 1 and 0 elsewhere. These two basis matrices are referred to as the exchangeable and the AR(1) working correlation matrices, respectively. The third one is the QREL estimator under the auxiliary mean model $y \sim x_1 + x_2$. The fourth one is the QREL2 estimator under the auxiliary mean model $y \sim x_1$. Note that under Setups 1 and 3, QREL uses a correct mean model while QREL2 is based on a misspecified mean model, so their performances will help us to understand the influence of the auxiliary mean structure to the QREL estimator. And similarly as in [26, Section 3], we consider AR(1) and exchangeable working correlation matrices

Table 2 Comparison of relative efficiency (RE). QREL, the mean model is $y \sim x_1 + x_2$, QREL2, the mean model is $y \sim x_1$; HFF, the estimator in [7, Section 2.2]; AR, autoregressive; EX, exchangeable

			$\alpha = 0.9$		$\alpha = 0.7$		$\alpha = 0.3$	
Setup	$V(\alpha)$	Method	β_1	β_2	β_1	β_2	β_1	β_2
1	EX	NEL	2.422	2.144	1.349	1.429	1.021	1.022
		AEL	2.343	2.024	1.314	1.341	0.996	0.993
		QREL	1.496	1.421	1.191	1.302	1.019	1.071
		QREL2	1.503	0.962	1.210	0.992	1.027	0.980
		HFF	2.511	2.271	1.417	1.541	0.975	1.019
	AR	NEL	2.355	2.571	1.467	1.352	1.064	1.041
		AEL	2.323	2.376	1.407	1.310	1.028	0.990
		QREL	1.459	1.549	1.380	1.287	1.059	1.037
		QREL2	1.463	0.976	1.364	0.971	1.073	0.981
		HFF	2.244	2.295	1.285	1.275	0.968	0.975
2	EX	NEL	2.298	2.240	1.594	1.393	1.024	1.002
		AEL	2.126	2.043	1.446	1.307	1.012	0.973
		QREL	1.135	0.955	1.009	1.069	0.923	0.923
		QREL2	1.097	0.932	1.007	0.949	0.966	0.937
		HFF	2.631	2.469	1.776	1.645	1.061	1.115
	AR	NEL	2.376	2.439	1.480	1.392	1.050	1.027
		AEL	2.252	2.292	1.450	1.354	1.021	1.022
		QREL	0.973	1.007	0.985	0.944	0.942	0.966
		QREL2	0.964	0.946	0.997	0.956	0.943	0.971
		HFF	2.264	2.300	1.271	1.222	0.999	1.048
3	EX	NEL	2.160	2.530	1.369	1.290	0.986	1.002
		AEL	2.122	2.367	1.330	1.257	0.954	0.974
		QREL	1.330	1.557	1.206	1.217	0.982	1.017
		QREL2	1.327	0.998	1.196	0.985	1.008	0.965
		HFF	2.371	2.489	1.551	1.492	1.066	1.020
	AR	NEL	2.181	2.282	1.396	1.474	1.055	1.019
		AEL	2.095	2.220	1.329	1.446	1.030	0.992
		QREL	1.364	1.391	1.176	1.273	1.005	0.982
		QREL2	1.377	0.979	1.179	0.999	1.014	0.981
		HFF	2.164	2.218	1.382	1.387	0.969	1.004

when computing the QREL estimators. The last one is the HFF estimator solves $\sum_{i=1}^N x_i W^{-1} S_i(\beta) = 0$ with $W = \frac{1}{N}\sum_{i=1}^N S_i(\hat{\beta}) S_i^{\mathrm{T}}(\hat{\beta})$. We use a stepwise replication of the bisection method to get the HFF estimator. Given $\beta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)})$, the new estimate $\beta_l^{(1)}$ of the l-th parameter ($l = 1, 2$) is obtained by applying the bisection method to the estimating equation for β_l, fixing the other argument at the current estimate. We cyclicly update the parameter until the estimates converge.

We compare the above five estimators in their relative efficiency (RE), defined as the ratio of the root mean squared error of the CQ estimator to that of the other estimators. A larger RE indicates better efficiency. We first use the exchangeable working correlation structure for both NEL and AEL estimators and the QREL estimators. See results in Table 1. We see that the NEL and AEL estimators perform similarly across different setups, and always outperform the CQ, QREL and QREL2 estimators under moderate to large latent correlation ($\alpha = 0.7$ or 0.9), even when using a misspecified working correlation structure, i.e., $V(\alpha) = $ AR. The advantage becomes more evident as the latent correlation α gets stronger. Moreover, we also see that the QREL estimators no longer provide much efficiency gain over the CQ estimator under Setups 2 and 3, when either the mean or variance does not exist. These observations suggest that our empirical likelihood weighting method is a more efficient way to utilize within-subject correlations than the method of [26]. Furthermore, we also notice that using a correct working correlation matrix improves the efficiency for the NEL and AEL estimators. The same fact applies to the QREL estimators as well. Results based on the AR(1) working correlation structure are similar and presented in Table 2.

Table 3 The estimated coverage probabilities and the mean lengths of confidence intervals (in parentheses) with 1,000 simulations in Setup 1. CQ is the conventional quantile regression

			$V(\alpha)$			
			EX		AR	
α	Working correlation	Method	β_1	β_2	β_1	β_2
0.9	–	CQ	0.953 (0.308)	0.957 (0.613)	0.955 (0.303)	0.954 (0.601)
	EX	NEL	0.950 (0.179)	0.953 (0.362)	0.949 (0.226)	0.953 (0.452)
		AEL	0.951 (0.178)	0.954 (0.360)	0.950 (0.225)	0.951 (0.451)
	AR	NEL	0.945 (0.209)	0.952 (0.421)	0.947 (0.198)	0.951 (0.395)
		AEL	0.935 (0.203)	0.942 (0.407)	0.940 (0.194)	0.945 (0.386)
0.7	–	CQ	0.954 (0.302)	0.953 (0.599)	0.953 (0.298)	0.949 (0.587)
	EX	NEL	0.946 (0.228)	0.950 (0.456)	0.949 (0.271)	0.948 (0.534)
		AEL	0.942 (0.228)	0.949 (0.455)	0.947 (0.270)	0.944 (0.532)
	AR	NEL	0.942 (0.253)	0.946 (0.504)	0.947 (0.244)	0.943 (0.481)
		AEL	0.938 (0.249)	0.939 (0.494)	0.939 (0.241)	0.937 (0.474)
0.3	–	CQ	0.952 (0.297)	0.953 (0.586)	0.953 (0.296)	0.948 (0.580)
	EX	NEL	0.944 (0.276)	0.950 (0.546)	0.947 (0.291)	0.944 (0.571)
		AEL	0.942 (0.275)	0.944 (0.543)	0.947 (0.289)	0.943 (0.566)
	AR	NEL	0.945 (0.286)	0.948 (0.566)	0.948 (0.285)	0.946 (0.558)
		AEL	0.938 (0.281)	0.937 (0.554)	0.942 (0.280)	0.938 (0.548)

Table 4 The estimated coverage probabilities and the mean lengths of confidence intervals (in parentheses) with 1,000 simulations in Setup 2. CQ is the conventional quantile regression

			$V(\alpha)$			
			EX		AR	
α	Working correlation	Method	β_1	β_2	β_1	β_2
0.9	–	CQ	0.954 (0.282)	0.958 (0.573)	0.946 (0.278)	0.955 (0.563)
	EX	NEL	0.945 (0.164)	0.949 (0.336)	0.948 (0.208)	0.949 (0.422)
		AEL	0.945 (0.165)	0.950 (0.339)	0.945 (0.209)	0.948 (0.425)
	AR	NEL	0.944 (0.192)	0.950 (0.391)	0.942 (0.182)	0.949 (0.369)
		AEL	0.931 (0.188)	0.935 (0.383)	0.936 (0.179)	0.941 (0.364)
0.7	–	CQ	0.950 (0.278)	0.950 (0.561)	0.944 (0.273)	0.947 (0.551)
	EX	NEL	0.943 (0.209)	0.945 (0.426)	0.936 (0.248)	0.942 (0.501)
		AEL	0.940 (0.210)	0.942 (0.428)	0.934 (0.248)	0.939 (0.501)
	AR	NEL	0.942 (0.232)	0.944 (0.470)	0.938 (0.223)	0.940 (0.450)
		AEL	0.930 (0.229)	0.936 (0.464)	0.932 (0.221)	0.931 (0.446)
0.3	–	CQ	0.947 (0.273)	0.944 (0.549)	0.944 (0.269)	0.944 (0.543)
	EX	AEL	0.940 (0.253)	0.937 (0.510)	0.934 (0.263)	0.936 (0.530)
		NEL	0.942 (0.253)	0.940 (0.511)	0.936 (0.265)	0.941 (0.534)
	AR	NEL	0.938 (0.262)	0.940 (0.529)	0.936 (0.259)	0.937 (0.521)
		AEL	0.932 (0.259)	0.933 (0.521)	0.928 (0.256)	0.929 (0.513)

Table 5 The estimated coverage probabilities and the mean lengths of confidence intervals (in parentheses) with 1,000 simulations in Setup 3. CQ is the conventional quantile regression

			$V(\alpha)$			
			EX		AR	
α	Working correlation	Method	β_1	β_2	β_1	β_2
0.9	–	CQ	0.956 (0.214)	0.958 (0.430)	0.949 (0.212)	0.953 (0.424)
	EX	NEL	0.947 (0.123)	0.947 (0.249)	0.946 (0.158)	0.945 (0.317)
		AEL	0.946 (0.123)	0.946 (0.250)	0.945 (0.158)	0.943 (0.317)
	AR	NEL	0.943 (0.144)	0.949 (0.290)	0.940 (0.137)	0.945 (0.275)
		AEL	0.930 (0.141)	0.942 (0.284)	0.934 (0.135)	0.938 (0.272)
0.7	–	CQ	0.946 (0.212)	0.951 (0.424)	0.947 (0.209)	0.949 (0.419)
	EX	NEL	0.942 (0.159)	0.946 (0.321)	0.942 (0.190)	0.944 (0.381)
		AEL	0.941 (0.160)	0.943 (0.322)	0.939 (0.190)	0.944 (0.381)
	AR	NEL	0.941 (0.177)	0.943 (0.355)	0.940 (0.171)	0.942 (0.343)
		AEL	0.934 (0.175)	0.938 (0.350)	0.934 (0.169)	0.936 (0.340)
0.3	–	CQ	0.944 (0.209)	0.944 (0.419)	0.943 (0.209)	0.947 (0.418)
	EX	NEL	0.938 (0.194)	0.938 (0.390)	0.939 (0.205)	0.943 (0.411)
		AEL	0.938 (0.194)	0.937 (0.389)	0.937 (0.204)	0.942 (0.409)
	AR	NEL	0.935 (0.201)	0.940 (0.404)	0.941 (0.201)	0.941 (0.402)
		AEL	0.931 (0.199)	0.934 (0.398)	0.934 (0.198)	0.935 (0.397)

An interesting observation is that the QREL2 estimator always shows efficiency gain over the CQ estimator on β_1 but not so on β_2. Recall that the auxiliary mean model used by QREL2 only contains covariate x_1 but not x_2, so this fact reflects that the correctness of the working mean model is critical to the effectiveness of QREL.

It is also interesting to see that the HFF estimator in general performs quite well, and sometimes beat our NEL and AEL estimators if we use a misspecified working correlation structure. So, it can be a good choice if the working correlation is likely misspecified. But it can only be used for balanced data and does not provide efficiency gain over the CQ estimator for large n and small N (see the Supplementary material of [26]).

Tables 3–5 summarize the estimated coverage probabilities and the average lengths of 95% confidence intervals of the CQ, NEL and AEL methods based on 1,000 simulations. The three methods all give coverage probabilities close to the nominal level of 95%. The NEL and AEL give similar mean lengths of confidence intervals and the mean lengths of confidence intervals of these two methods are smaller than those of the CQ method under moderate to large latent correlation ($\alpha = 0.7$ or 0.9).

5 Data analysis

We analyzed the CD4 data in [3], where 2,376 CD4 observations were obtained from 369 infected men enrolled in the Multicenter AIDS Cohort Study (see [10]). Time is measured in years with the origin at the date of seroconversion. We model the population quantiles of the square-root-transformed CD4 counts as a function of time and the following additional covariates: age at the time of seroconversion, packs of cigarettes smoked per day, recreational drug use, number of sexual partners and the depressive symptoms, where severity of depressive symptoms was evaluated with the Center for Epidemiologic Studies Depression Scale (CES-D Scale). The transformation on CD4 counts was suggested in [3] to achieve better normality. We model the time trend as a constant prior to seroconversion and quadratic in time thereafter.

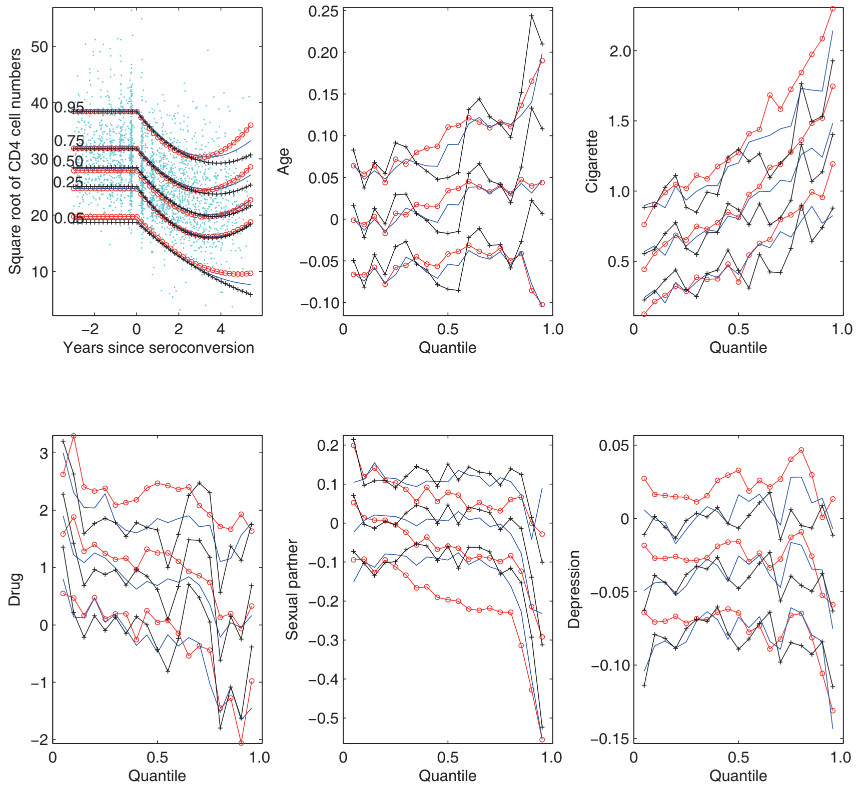

Figure 1 The top left panel depicts the data points and the fitted quantile curves for time at $\tau = 0.05, 0.25, 0.50, 0.75$ and 0.95, using our NEL method $(-)$, AEL method $(-+)$ and the conventional quantile regression method $(-o)$. The rest are fitted quantile coefficients for the NEL method $(-)$ with 95% confidence intervals $(-)$; for the AEL method $(-+)$ with 95% confidence intervals $(-+)$; for the conventional quantile regression method $(-o)$ with 95% confidence intervals $(-o)$

We compare three estimators: the CQ estimator that ignores the within-subject correlations, the NEL and AEL estimators. For the NEL and AEL estimators, we base $g(z_i, \beta)$ and $\tilde{g}(z_i, \beta)$ on the AR(1) working correlation matrix. The confidence regions for these three approaches are conducted based on the normal approximation. In the first panel of Figure 1, we plot the data with the estimated time trends at different quantiles for these three methods. The three methods produce similar trends when $\tau = 0.25, 0.5$ and $\tau = 0.75$. However, when $\tau = 0.05$ and 0.95, there is a considerable difference between the estimated time trends. We plot the estimated regression coefficients for the five main convariates at various quantile levels. We see that the NEL and AEL methods produce similar estimated regression coefficients for the five main convariates at various quantile levels. The differences between the CQ estimator and AEL estimator of the regression coefficients exist for drug and sexual partners.

6 Comments and conclusions

In this paper, we develop weighted quantile regression estimators for longitudinal data. The weights are optimized by empirical likelihood with estimating equation that incorporates within-subject correlation, and hence achieve more efficient estimation. We show that the proposed estimators are asymptotically normal and more efficient than the conventional quantile regression estimator. Our approach is partly motivated by the QREL estimator in [26]. Compared with QREL, our proposed estimator does not require specification of the auxiliary mean structure, and display further efficiency gain in simulation studies.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 11401048, 11301037, 11571051 and 11201174) and the Natural Science Foundation for Young Scientists of Jilin Province of China (Grant Nos. 20150520055JH and 20150520054JH). The authors thank anonymous referees for their careful reading of the paper and insightful comments.

References

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

2 Chen L, Wei L J, Parzen M I. Quantile regression for correlated observations. In: Proceedings of the Second Seattle Symposium in Biostatistics. Lecture Notes in Statistics, vol. 179. New York: Springer, 2004, 51–69

3 Diggle P J, Heagerty P, Liang K Y, et al. Analysis of Longitudinal Data, 2nd ed. New York: Oxford University Press, 2002

4 Farcomeni A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. Stat Comput, 2012, 22: 141–152

5 Fu L, Wang Y. Quantile regression for longitudinal data with a working correlation model. Comp Statist Data Anal, 2012, 56: 2526–2538

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

7 He X, Fu B, Fung W. Median regression for longitudinal data. Statist Med, 2003, 22: 3655–3669

8 He X, Shao Q. A general Bahadur representation of M-estimators and its application to linear regression with non-stochastic designs. Ann Statist, 1996, 24: 2608–2630

9 Jung S H. Quasi-likelihood for median regression models. J Amer Statist Assoc, 1996, 91: 251–257

10 Kaslow R A, Ostrow D C, Detels R, et al. The multicenter AIDS cohort study: Rationale, organization, and selected characteristics of the participants. Amer J Epidemiol, 1987, 126: 310–318

11 Kitamura Y, Tripathi G, Ahn H. Empirical likelihood-based inference in conditional moment restriction models. Econometrica, 2004, 72: 1667–1714

12 Knight K. Limiting distributions for L_1 regression estimators under general conditions. Ann Statist, 1998, 26: 755–770

13 Koenker R. Quantile regression for longitudinal data. J Mult Anal, 2004, 91: 74–89

14 Koenker R. Quantile Regression. Cambridge: Cambridge University Press, 2005

15 Koenker R, Bassett G. Regression quantiles. Econometrica, 1978, 46: 33–50

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

17 Liu Y, Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. Int J Biostat, 2009, 5: Article 28

18 Mu Y, He X. Power transformation toward a linear regression quantile. J Amer Statist Assoc, 2007, 102: 269–279

19 Owen A. Empirical likelihood ratio confidence regions. Ann Statist, 1990, 18: 90–120

20 Owen A. Empirical Likelihood. New York: Chapman and Hall-CRC, 2001

21 Pollard D. Empirical Processes: Theory and Applications. Hayward: Institute of Mathematical Statistics, 1990

22 Qu A, Lindsay B G, Li B. Improving generalised estimating equations using quadratic inference functions. Biometrika, 2000, 87: 823–836

23 Reich B J, Bondell H D, Wang H J. Flexible Bayesian quantile regression for independent and clustered data. Biostatistics, 2010, 11: 337–352

24 Robins J M, Rotnizky A, Zhao L P. Estimation of regression coefficients when some regressors are not always observed. J Amer Statist Assoc, 1994, 89: 846–886

25 Robins J M, Rotnitzky A, Zhao L P. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J Amer Statist Assoc, 1995, 90: 106–121

26 Tang C Y, Leng C. Empirical likelihood and quantile regression in longitudinal data analysis. Biometrika, 2011, 98: 1001–1006

27 Wang H J, Zhu Z. Empirical likelihood for quantile regression models with longitudinal data. J Statist Plan Infer, 2011, 141: 1603–1615

28 Yin G, Cai J. Quantile regression models with multivariate failure time data. Biometrics, 2005, 61: 151–161

29 Yu K, Zhang J. A three-parameter asymmetric Laplace distribution and its extension. Commun Statist A, 2005, 34: 1867–1879

Appendix

C1. $z_i = (y_i, x_i^{\mathrm{T}})$, $i = 1, \ldots, N$, are independent.

C2. The true parameter value β_0 is an interior point of a compact parameter space $\Theta \subset \mathbb{R}^p$.

C3. The τ-th conditional quantile of y_{it} given x_i is given by $Q_\tau(y_{it} \mid x_i) = x_{it}^{\mathrm{T}}\beta_0$.

C4. $n^* = \sup_i n_i < \infty$ and x_i has a bounded support.

C5. Let $F_{it}(\cdot)$ and $f_{it}(\cdot)$ denote respectively the conditional distribution and density functions of $\varepsilon_{it} = y_{it} - x_{it}^{\mathrm{T}}\beta_0$ given x_i. The distribution functions $F_{it}(\cdot)$'s are absolutely continuous, with continuous densities

$f_{it}(\cdot)$'s that are uniformly bounded away form 0 and ∞ at 0. Also, $f'_{it}(\cdot)$ exists and is uniformly bounded.
C6. D_0, D_1, Π and Δ are positive definite.

Lemma A.1. *If* $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E\{\sup_{\beta\in\Theta}\|g(z_i,\beta)\|^q\} < \infty$, *then* $\max_{1\leqslant i\leqslant N}\sup_{\beta\in\Theta}\|g(z_i,\beta)\| = o(N^{1/q})$ *almost surely.*

Proof. See the proof of [11, Lemma D.2]. □

Lemma A.2. *For some* $0 < C < \infty$,

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}}\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)-\frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0)-D_2(\beta-\beta_0)\right\| = o_p(N^{-1/2}),$$

where $D_2=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N(I_m\otimes x_i)M_i\Psi_i x_i^{\mathrm{T}}$ *and* $\Psi_i=\mathrm{diag}\{f_{i1}(0),\ldots,f_{in_i}(0)\}$.

Proof. Let

$$\bar g(z_i,\beta)=(I_m\otimes x_i)M_i\bar S_i(\beta),\quad i=1,\ldots,N,$$

where $\bar S_i(\beta)=(\bar S_{i1}(\beta),\ldots,\bar S_{in_i}(\beta))^{\mathrm{T}}$ and $\bar S_{it}(\beta)=P(y_{it}-x_{it}^{\mathrm{T}}\beta\leqslant 0)-\tau$. Note that

$$\begin{aligned}&\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)-\frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0)-D_2(\beta-\beta_0)\right\|\\ &\quad\leqslant\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)-\frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0)-\frac{1}{N}\sum_{i=1}^N\bar g(z_i,\beta)\right\|\\ &\qquad+\left\|\frac{1}{N}\sum_{i=1}^N\bar g(z_i,\beta)-D_2(\beta-\beta_0)\right\|.\end{aligned}$$

According to the lemma in [9], the first term

$$\sup_{|\beta-\beta_0|<CN^{-1/2}}\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)-\frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0)-\frac{1}{N}\sum_{i=1}^N\bar g(z_i,\beta)\right\| = o_p(N^{-1/2}).$$

The second term

$$\begin{aligned}&\sup_{|\beta-\beta_0|<CN^{-1/2}}\left\|\frac{1}{N}\sum_{i=1}^N\bar g(z_i,\beta)-D_2(\beta-\beta_0)\right\|\\ &\quad=\left\|\frac{1}{N}\sum_{i=1}^N(I_m\otimes x_i)M_i\Psi_i x_i^{\mathrm{T}}-D_2\right\|\sup_{|\beta-\beta_0|<CN^{-1/2}}\|\beta-\beta_0\|+o_p(N^{-1/2}).\end{aligned}$$

It follows from the law of large numbers that $\|N^{-1}\sum_{i=1}^N(I_m\otimes x_i)M_i\Psi_i x_i^{\mathrm{T}}-D_2\|=o_p(1)$. This completes the proof. □

Lemma A.3. *For some* $0 < C < \infty$,

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}}\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)-\Pi\right\| = o_p(1),$$

where $\Pi=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\{(I_m\otimes x_i)M_i\}\Lambda_i\{(I_m\otimes x_i)M_i\}^{\mathrm{T}}$ *and* $\Lambda_i=E\{S_i(\beta_0)S_i^{\mathrm{T}}(\beta_0)\mid x_i\}$.

Proof. Let $\tilde\Pi(\beta)=\frac{1}{N}\sum_{i=1}^N E[\{(I_m\otimes x_i)M_i\}S_i(\beta)S_i^{\mathrm{T}}(\beta)\{(I_m\otimes x_i)M_i\}^{\mathrm{T}}]$. By the triangle inequality,

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}}\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)-\Pi\right\|$$

$$\leqslant \sup_{\|\beta-\beta_0\|<CN^{-1/2}} \left\| \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta) - \tilde{\Pi}(\beta) \right\| + \sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\tilde{\Pi}(\beta)-\tilde{\Pi}(\beta_0)\| + \|\tilde{\Pi}(\beta_0)-\Pi\|.$$

Under Condition C4, by using the uniform strong law of large numbers (see [21, p. 41]), we have

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \left\| \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta) - \tilde{\Pi}(\beta) \right\| = o(1) \ \text{ almost surely.}$$

It is easy to see that $\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\tilde{\Pi}(\beta)-\tilde{\Pi}(\beta_0)\| = o_p(1)$ and $\|\tilde{\Pi}(\beta_0)-\Pi\| = o_p(1)$. The desired result follows. □

Lemma A.4. *If $\lambda=\lambda(\beta)$ solves*

$$\sum_{i=1}^N \frac{g(z_i,\beta)}{1+\lambda^{\mathrm{T}}g(z_i,\beta)} = 0, \tag{A.1}$$

then we have $\|\lambda(\beta)\| = O_p(N^{-1/2})$ and

$$\lambda(\beta) = \left\{ \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta) \right\}^{-1} \frac{1}{N}\sum_{i=1}^N g(z_i,\beta) + o_p(N^{-1/2})$$

uniformly about $\beta\in B_0=\{\beta:\|\beta-\beta_0\|<CN^{-1/2}\}$ for some $0<C<\infty$.

Proof. The basic idea behind this proof is outlined in [19].

Let $U_i=\lambda^{\mathrm{T}}g(z_i,\beta)$ and $g^*=\sup_{\beta\in B_0}\max_{1\leqslant i\leqslant N}\|g(z_i,\beta)\|$. Let $\lambda(\beta)=\|\lambda(\beta)\|v$, $\|v\|=1$. Substituting $1/(1+U_i)=1-U_i/(1+U_i)$ into (A.1) and simplifying, we find that

$$\|\lambda(\beta)\| v^{\mathrm{T}} \frac{1}{N}\sum_{i=1}^N \frac{g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)}{1+U_i} v = v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta).$$

Since every $p_i>0$, we have $1+U_i>0$ and therefore

$$\begin{aligned}\|\lambda(\beta)\| v^{\mathrm{T}} \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)v &\leqslant \|\lambda(\beta)\| v^{\mathrm{T}} \frac{1}{N}\sum_{i=1}^N \frac{g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)}{1+U_i} v(1+\|\lambda(\beta)\|g^*) \\ &= v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)(1+\|\lambda(\beta)\|g^*).\end{aligned}$$

Consequently,

$$\|\lambda(\beta)\| \left(v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)v - g^* v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta) \right) \leqslant v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta).$$

By Lemma A.2,

$$\begin{aligned}\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \left\| \frac{1}{N}\sum_{i=1}^N g(z_i,\beta) \right\| &\leqslant \sup_{\|\beta-\beta_0\|<CN^{-1/2}} \left\| \frac{1}{N}\sum_{i=1}^N g(z_i,\beta) - \frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0) - D_2(\beta-\beta_0) \right\| \\ &\quad + \left\| \frac{1}{N}\sum_{i=1}^N g(z_i,\beta_0) \right\| + C\|D_2\|N^{-1/2} \\ &= O_p(N^{-1/2}).\end{aligned}$$

Furthermore, by Lemma A.1, we have $g^*=o_p(N^{1/2})$. Thus we have

$$\|\lambda(\beta)\| \left(v^{\mathrm{T}}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)v + o_p(1) \right) = O_p(N^{-1/2}).$$

Then Lemma A.3 gives $\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)\| = O_p(1)$. Hence, uniformly for any $\beta \in \{\beta \mid \|\beta-\beta_0\| < CN^{-1/2}\}$, we obtain $\|\lambda(\beta)\| = O_p(N^{-1/2})$. Based on this order bound for $\lambda(\beta)$, we then have $\max_{1\leqslant i\leqslant N}\|U_i\| = o_p(1)$ uniformly for $\beta \in \{\beta \mid \|\beta-\beta_0\| < CN^{-1/2}\}$.

Now, write

$$\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)/(1+U_i)\\
&= \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)\left(1-U_i+\frac{U_i^2}{1+U_i}\right)\\
&= \frac{1}{N}\sum_{i=1}^N g(z_i,\beta) - \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)\lambda + \frac{1}{N}\sum_{i=1}^N g(z_i,\beta)\frac{U_i^2}{1+U_i}.
\end{aligned}$$

Because

$$\begin{aligned}
\left\|\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)\frac{U_i^2}{1+U_i}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^N \|g(z_i,\beta)\|^3\|\lambda\|^2|1+U_i|^{-1}\\
&= o_p(N^{1/2})O_p(N^{-1})O_p(1) = o_p(N^{-1/2}),
\end{aligned}$$

we have, uniformly for $\beta \in \{\beta \mid \|\beta-\beta_0\| < CN^{-1/2}\}$,

$$\lambda(\beta) = \left\{\frac{1}{N}\sum_{i=1}^N g(z_i,\beta)g^{\mathrm{T}}(z_i,\beta)\right\}^{-1}\frac{1}{N}\sum_{i=1}^N g(z_i,\beta) + o_p(N^{-1/2}). \qquad \square$$

Lemma A.5. *Let $\varepsilon_{it}(\beta) = y_{it} - x_{it}^{\mathrm{T}}\beta$, $\sigma_{it}^2 = x_{it}^{\mathrm{T}}x_{it}/N$, $\hat{\Psi}_{it}(\beta) = \sigma_{it}^{-1}\phi(\varepsilon_{it}(\beta)/\sigma_{it})$ and $\hat{\Psi}_i(\beta) = \mathrm{diag}\{\hat{\Psi}_{i1}(\beta),\ldots,\hat{\Psi}_{in_i}(\beta)\}$, where $\phi(u) = (2\pi)^{-1/2}\exp(-u^2/2)$. Let $\hat{D}_1(\beta) = \frac{1}{N}\sum_{i=1}^N x_i\hat{\Psi}_i(\beta)x_i^{\mathrm{T}}$ and $\hat{D}_2(\beta) = \frac{1}{N}\sum_{i=1}^N (I_m\otimes x_i)M_i\hat{\Psi}_i(\beta)x_i^{\mathrm{T}}$. Then, we have*

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\hat{D}_1(\beta) - D_1\| \xrightarrow{p} 0, \tag{A.2}$$

$$\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\hat{D}_2(\beta) - D_2\| \xrightarrow{p} 0. \tag{A.3}$$

Proof. The two results (A.2) and (A.3) can be proved in the same way, so we only prove (A.2) here. By the triangle inequality, we have

$$\begin{aligned}
\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\hat{D}_1(\beta) - D_1\| \leqslant{}& \sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\hat{D}_1(\beta) - \hat{D}_1(\beta_0)\| + \|\hat{D}_1(\beta_0) - E\{\hat{D}_1(\beta_0) \mid x_1,\ldots,x_N\}\|\\
&+ \|E\{\hat{D}_1(\beta_0) \mid x_1,\ldots,x_N\} - \tilde{D}_1\| + \|\tilde{D}_1 - D_1\|,
\end{aligned}$$

where $\tilde{D}_1 = \frac{1}{N}\sum_{i=1}^N x_i\Psi_i x_i^{\mathrm{T}}$. It is easy to see that $\|\tilde{D}_1 - D_1\| \xrightarrow{p} 0$ and

$$\begin{aligned}
\sup_{\|\beta-\beta_0\|<CN^{-1/2}} \|\hat{D}_1(\beta) - \hat{D}_1(\beta_0)\| &\leqslant \sup_{\|\beta-\beta_0\|<CN^{-1/2}} \frac{1}{N}\sum_{i=1}^N\sum_{t=1}^{n_i} \|x_{it}\|^2|\hat{\Psi}_{it}(\beta) - \hat{\Psi}_{it}(\beta_0)|\\
&\leqslant CN^{-1/2}\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^{n_i}\|x_{it}\|^3\sigma_{it}^{-2}|\phi'(\hat{\varepsilon}_{it}(\beta^*)/\sigma_{it})|,
\end{aligned}$$

where $\phi'(u)$ is the derivative of $\phi(u)$ and β^* is a point on the segment connecting β and β_0. Note that $\sigma_{it} = O_p(N^{-1/2})$ and $\lim_{u\to\infty}|u\phi'(u)| = 0$, then we have $\sup_{\|\beta-\beta_0\|<CN^{-1/2}}\|\hat{D}_1(\beta) - \hat{D}_1(\beta_0)\| = o_p(1)$. By the law of large numbers, we have $\|\hat{D}_1(\beta_0) - E\{\hat{D}_1(\beta_0) \mid x_1,\ldots,x_N\}\| \xrightarrow{p} 0$. Next, we will show that $\|E\{\hat{D}_1(\beta_0) \mid x_1,\ldots,x_N\} - \tilde{D}_1\| \xrightarrow{p} 0$. Observe that

$$\|E\{\hat{D}_1(\beta_0) \mid x_1,\ldots,x_N\} - \tilde{D}_1\| = \left\|\frac{1}{N}\sum_{i=1}^N x_i[E\{\hat{\Psi}_i(\beta_0) \mid x_i\} - \Psi_i]x_i^{\mathrm{T}}\right\|$$

$$\leqslant \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{n_i}\|x_{it}\|^2|E\{\hat{\Psi}_{it}(\beta_0)\mid x_{it}\}-\Psi_{it}|.$$

Furthermore,

$$\begin{aligned}|E\{\hat{\Psi}_{it}(\beta_0)\mid x_{it}\}-\Psi_{it}| &= \left|\sigma_{it}^{-1}\int \phi(u/\sigma_{it})f_{it}(u)du - f_{it}(0)\right|\\ &= \left|\int \phi(u)[f_{it}(0)+u\sigma_{it}f'_{it}(\xi_u)]du - f_{it}(0)\right|\\ &\leqslant \sigma_{it}\int|\phi(u)uf'_{it}(\xi_u)|du,\end{aligned}$$

where ξ_u lies between 0 and $u\sigma_{it}$. By the condition C5, there exists $M>0$, such that $|f'_{it}(u)|<M$. It follows that

$$|E\{\hat{\Psi}_{it}(\beta_0)\mid x_{it}\}-\Psi_{it}|\leqslant \sigma_{it}\sqrt{\frac{2}{\pi}}M\to 0.$$

Thus, we have $\|E\{\hat{D}_1(\beta_0)\mid x_1,\dots,x_N\}-\tilde{D}_1\|\to 0$. The desired result follows. □

Proof of Theorem 2.1. See the proof of [27, Theorem 1]. □

Proof of Theorem 3.1. By the proof of Theorem 2.1, we have $\sqrt{N}(\hat{\beta}-\beta_0)=-D_1^{-1}B+o_p(1)$, where $D_1=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_i\Psi_i x_i^{\mathrm{T}}$, $B=N^{-1/2}\sum_{i=1}^{N}x_iS_i(\beta_0)\xrightarrow{d}N(0,D_0)$, $D_0=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_i\Lambda_i x_i^{\mathrm{T}}$ and $\Lambda_i=E\{S_i(\beta_0)S_i^{\mathrm{T}}(\beta_0)\mid x_i\}$. By Lemma A.2,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}g(z_i,\hat{\beta}) &= \frac{1}{N}\sum_{i=1}^{N}g(z_i,\beta_0)+D_2(\hat{\beta}-\beta_0)+o_p(N^{-1/2})\\ &= \frac{1}{N}\sum_{i=1}^{N}g(z_i,\beta_0)-D_2D_1^{-1}N^{-1/2}B+o_p(N^{-1/2}).\end{aligned}$$

Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}g(z_i,\hat{\beta})=\frac{1}{N}\sum_{i=1}^{N}\{(I_m\otimes x_i)M_i-D_2D_1^{-1}x_i\}S_i(\beta_0)+o_p(N^{-1/2}).$$

By Lemmas A.3 and A.4, we have

$$\begin{aligned}\hat{\lambda} &= \left\{\frac{1}{N}\sum_{i=1}^{N}g(z_i,\hat{\beta})g(z_i,\hat{\beta})^{\mathrm{T}}\right\}^{-1}\frac{1}{N}\sum_{i=1}^{N}g(z_i,\hat{\beta})+o_p(N^{-1/2})\\ &= \Pi^{-1}\frac{1}{N}\sum_{i=1}^{N}\{(I_m\otimes x_i)M_i-D_2D_1^{-1}x_i\}S_i(\beta_0)+o_p(N^{-1/2}).\end{aligned}\tag{A.4}$$

Let $\varepsilon_{it}=y_{it}-x_{it}^{\mathrm{T}}\beta_0$ and $A_{it}(\delta)=\rho_\tau(\varepsilon_{it}-x_{it}^{\mathrm{T}}\delta/\sqrt{N})-\rho_\tau(\varepsilon_{it})$. The function $A_N(\delta)=\sum_{i=1}^{N}N\hat{p}_i\sum_{t=1}^{n_i}A_{it}(\delta)$ is convex and is minimized at $\hat{\delta}=\sqrt{N}(\hat{\beta}_{\mathrm{NEL}}-\beta_0)$. Utilizing the following identity (see [12]),

$$\rho_\tau(u-v)-\rho_\tau(u)=v\psi_\tau(u)+\int_0^v\{I(u\leqslant s)-I(u\leqslant 0)\}ds,$$

where $\psi_\tau(u)=I(u<0)-\tau$, we have $A_{it}(\delta)=A_{1it}(\delta)+A_{2it}(\delta)$ with

$$\begin{aligned}A_{1it}(\delta) &= N^{-1/2}x_{it}^{\mathrm{T}}\delta S_{it}(\beta_0),\\ A_{2it}(\delta) &= \int_0^{x_{it}^{\mathrm{T}}\delta/\sqrt{N}}\{I(\varepsilon_{it}\leqslant s)-I(\varepsilon_{it}\leqslant 0)\}ds.\end{aligned}$$

Thus,

$$A_N(\delta) = \sum_{i=1}^{N} N\hat{p}_i \sum_{t=1}^{n_i} A_{1it}(\delta) + \sum_{i=1}^{N} N\hat{p}_i \sum_{t=1}^{n_i} A_{2it}(\delta).$$

Since $\hat{p}_i = 1/[N(1+\hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta}))]$, by Lemmas A.1 and A.4, we have that

$$\hat{p}_i = N^{-1}\{1 - \hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta})(1+o_p(1))\},$$

uniformly for $i = 1,\dots,N$. For the first term, we have

$$\begin{aligned}
&\sum_{i=1}^{N} N\hat{p}_i \bigg(\sum_{t=1}^{n_i} A_{1it}(\delta)\bigg) \\
&\quad = \sum_{i=1}^{N} \{1 - \hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta})(1+o_p(1))\} \bigg(\sum_{t=1}^{n_i} A_{1it}(\delta)\bigg) \\
&\quad = N^{-1/2} \sum_{i=1}^{N} \delta^{\mathrm{T}} x_i S_i(\beta_0) \{1 - \hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta})(1+o_p(1))\} \\
&\quad = \delta^{\mathrm{T}} N^{-1/2} \sum_{i=1}^{N} x_i S_i(\beta_0) - \delta^{\mathrm{T}} N^{-1} \sum_{i=1}^{N} x_i S_i(\beta_0) S_i^{\mathrm{T}}(\beta_0) M_i^{\mathrm{T}} (I_m \otimes x_i)^{\mathrm{T}} N^{1/2} \hat{\lambda} + o_p(1).
\end{aligned}$$

By the law of large numbers, we have that

$$N^{-1} \sum_{i=1}^{N} x_i S_i(\beta_0) S_i^{\mathrm{T}}(\beta_0) M_i^{\mathrm{T}} (I_m \otimes x_i)^{\mathrm{T}} \xrightarrow{p} \Gamma_1. \tag{A.5}$$

Thus, by (A.4) and (A.5), we obtain

$$\begin{aligned}
&\sum_{i=1}^{N} N\hat{p}_i \bigg(\sum_{t=1}^{n_i} A_{1it}(\delta)\bigg) \\
&\quad = \delta^{\mathrm{T}} N^{-1/2} \sum_{i=1}^{N} [x_i - \Gamma_1 \Pi^{-1}\{(I_m \otimes x_i) M_i - D_2 D_1^{-1} x_i\}] S_i(\beta_0) + o_p(1),
\end{aligned}$$

where $N^{-1/2}\sum_{i=1}^{N}[x_i - \Gamma_1\Pi^{-1}\{(I_m\otimes x_i)M_i - D_2 D_1^{-1}x_i\}]S_i(\beta_0) \xrightarrow{d} N(0, D_0 - \Omega_1)$ and Ω_1 is given in Theorem 3.1. For the second term, we have

$$\sum_{i=1}^{N} N\hat{p}_i \bigg(\sum_{t=1}^{n_i} A_{2it}(\delta)\bigg) = \sum_{i=1}^{N}\sum_{t=1}^{n_i} A_{2it}(\delta) - \sum_{i=1}^{N}\sum_{t=1}^{n_i} A_{2it}(\delta)\hat{\lambda}^{\mathrm{T}} g(z_i,\hat{\beta}).$$

We first show that

$$\sum_{i=1}^{N}\sum_{t=1}^{n_i} A_{2it}(\delta) = \frac{1}{2}\delta^{\mathrm{T}} D_1 \delta + o_p(1). \tag{A.6}$$

Observe that

$$\begin{aligned}
&E\bigg(\sum_{i=1}^{N}\sum_{t=1}^{n_i} A_{2it}(\delta) \,\bigg|\, x_1,\dots,x_N\bigg) \\
&\quad = \sum_{i=1}^{N}\sum_{t=1}^{n_i} E\{A_{2it}(\delta) \mid x_i\} \\
&\quad = \sum_{i=1}^{N}\sum_{t=1}^{n_i} \int_0^{x_{it}^{\mathrm{T}}\delta/\sqrt{N}} \{F_{it}(s) - F_{it}(0)\} ds
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \int_0^{x_{it}^{\mathrm{T}}\delta} \{F_{it}(u/\sqrt{N}) - F_{it}(0)\} du \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \int_0^{x_{it}^{\mathrm{T}}\delta} \sqrt{N}\{F_{it}(u/\sqrt{N}) - F_{it}(0)\} du \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \int_0^{x_{it}^{\mathrm{T}}\delta} f_{it}(0) u du + o_p(1) \\
&= (2N)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \delta^{\mathrm{T}} f_{it}(0) x_{it} x_{it}^{\mathrm{T}} \delta + o_p(1) \\
&= \frac{1}{2} \delta^{\mathrm{T}} D_1 \delta + o_p(1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{var}\Bigg(\sum_{i=1}^{N} \sum_{t=1}^{n_i} A_{2it}(\delta) \,\Bigg|\, x_1, \ldots, x_N \Bigg) \\
&= \sum_{i=1}^{N} \mathrm{var}\Bigg(\sum_{t=1}^{n_i} A_{2it}(\delta) \,\Bigg|\, x_i \Bigg) \\
&\leqslant \sum_{i=1}^{N} n_i^2 E\Bigg(\Bigg\{ n_i^{-1} \sum_{t=1}^{n_i} A_{2it}(\delta) \Bigg\}^2 \,\Bigg|\, x_i \Bigg) \\
&\leqslant \sum_{i=1}^{N} n_i \sum_{t=1}^{n_i} E(A_{2it}^2(\delta) \,|\, x_i) \\
&\leqslant \frac{n^*}{\sqrt{N}} \max_{1 \leqslant i \leqslant N, 1 \leqslant t \leqslant n_i} |x_{it}^{\mathrm{T}}\delta| \sum_{i=1}^{N} \sum_{t=1}^{n_i} E\{A_{2it}(\delta) \mid x_i\} = o_p(1),
\end{aligned}
$$

where $n^* = \sup_i n_i < \infty$. Therefore, (A.6) is proved. In addition, it is true that

$$
\Bigg| \sum_{i=1}^{N} \sum_{t=1}^{n_i} \hat{\lambda}^{\mathrm{T}} g(z_i, \hat{\beta}) A_{2it}(\delta) \Bigg| \leqslant \max_{1 \leqslant i \leqslant N} \{ |\hat{\lambda}^{\mathrm{T}} g(z_i, \hat{\beta})| \} \Bigg| \sum_{i=1}^{N} \sum_{t=1}^{n_i} A_{2it}(\delta) \Bigg|.
$$

Since $\max_{1 \leqslant i \leqslant N} \{ |\hat{\lambda}^{\mathrm{T}} g(z_i, \hat{\beta})| \} = o_p(1)$, this term is negligible asymptotically. It follows that

$$
\begin{aligned}
A_N(\delta) &= \sum_{i=1}^{N} N \hat{p}_i \Bigg(\sum_{t=1}^{n_i} A_{it}(\delta) \Bigg) \\
&\xrightarrow{d} A_0(\delta) = \delta^{\mathrm{T}} N^{-1/2} \sum_{i=1}^{N} \{ x_i - \Gamma_1 \Pi^{-1} [(I_m \otimes x_i) M_i - D_2 D_1^{-1} x_i] \} S_i(\beta_0) + \frac{1}{2} \delta^{\mathrm{T}} D_1 \delta.
\end{aligned}
$$

Summarizing the above results, we conclude that

$$
\sqrt{N}(\hat{\beta}_{\mathrm{NEL}} - \beta_0) = \hat{\delta} \xrightarrow{d} \arg\min_{\delta} A_0(\delta).
$$

The theorem follows by noting that

$$
\arg\min_{\delta} A_0(\delta) = -D_1^{-1} N^{-1/2} \sum_{i=1}^{N} \{ x_i - \Gamma_1 \Pi^{-1} [(I_m \otimes x_i) M_i - D_2 D_1^{-1} x_i] \} S_i(\beta_0). \qquad \square
$$

Proof of Theorem 3.2. By the proof of Theorem 3.1 and Lemma A.5, we have

$$
\frac{1}{N} \sum_{i=1}^{N} \tilde{g}(z_i, \hat{\beta}) = \frac{1}{N} \sum_{i=1}^{N} H_i S_i(\beta_0) + o_p(N^{-1/2}),
$$

where $H_i = (I_m \otimes x_i)M_i - D_2 D_1^{-1} x_i$. Similar to the proof of Lemmas A.3 and A.4, we have

$$\begin{aligned}\hat{\lambda} &= \left\{ \frac{1}{N}\sum_{i=1}^{N} \tilde{g}(z_i,\hat{\beta})\tilde{g}(z_i,\hat{\beta})^{\mathrm{T}} \right\}^{-1} \frac{1}{N}\sum_{i=1}^{N} \tilde{g}(z_i,\hat{\beta}) + o_p(N^{-1/2}) \\ &= \Delta^{-1}\frac{1}{N}\sum_{i=1}^{N} H_i S_i(\beta_0) + o_p(N^{-1/2}).\end{aligned}$$

Let $\varepsilon_{it} = y_{it} - x_{it}^{\mathrm{T}}\beta_0$ and $A_{it}(\delta) = \rho_\tau(\varepsilon_{it} - x_{it}^{\mathrm{T}}\delta/\sqrt{N}) - \rho_\tau(\varepsilon_{it})$. The function $A_N(\delta) = \sum_{i=1}^{N} N\hat{q}_i \sum_{t=1}^{n_i} A_{it}(\delta)$ is convex and is minimized at $\hat{\delta} = \sqrt{N}(\hat{\beta}_{\mathrm{AEL}} - \beta_0)$. By the similar arguments of the proof of Theorem 3.1, we obtain

$$\begin{aligned}A_N(\delta) &= \sum_{i=1}^{N} N\hat{q}_i \left(\sum_{t=1}^{n_i} A_{it}(\delta) \right) \\ &\xrightarrow{d} A_0(\delta) = \delta^{\mathrm{T}} N^{-1/2} \sum_{i=1}^{N} (x_i - \Gamma_2 \Delta^{-1} H_i) S_i(\beta_0) + \frac{1}{2}\delta^{\mathrm{T}} D_1 \delta,\end{aligned}$$

where $N^{-1/2}\sum_{i=1}^{N}(x_i - \Gamma_2\Delta^{-1}H_i)S_i(\beta_0) \xrightarrow{d} N(0, D_0 - \Omega_2)$ and Ω_2 is given in Theorem 3.2. Summarizing the above results, we conclude that

$$\sqrt{N}(\hat{\beta}_{\mathrm{AEL}} - \beta_0) = \hat{\delta} \xrightarrow{d} \arg\min_{\delta} A_0(\delta).$$

The theorem follows by noting that

$$\arg\min_{\delta} A_0(\delta) = -D_1^{-1} N^{-1/2} \sum_{i=1}^{N} (x_i - \Gamma_2 \Delta^{-1} H_i) S_i(\beta_0). \qquad \square$$