Empirical likelihood for quantile regression models with longitudinal data

Huixia Judy Wang, Zhongyi Zhu

Abstract

We develop two empirical likelihood-based inference procedures for longitudinal data under the framework of quantile regression. The proposed methods avoid estimating the unknown error density function and the intra-subject correlation involved in the asymptotic covariance matrix of the quantile estimators. By appropriately smoothing the quantile score function, the empirical likelihood approach is shown to have a higher-order accuracy through the Bartlett correction. The proposed methods exhibit finite-sample advantages over the normal approximation-based and bootstrap methods in a simulation study and the analysis of a longitudinal ophthalmology data set.

Keywords: Bartlett correction, Confidence region, Estimating equation, Hypothesis test, Kernel smoothing, Quantile regression

1. Introduction

Quantile regression has emerged as a powerful complement to the least squares regression, partly due to its flexibility and ability of describing the entire conditional distribution of a response variable. In the past three decades, estimation and asymptotic theory for various quantile regression models have been extensively studied; see Koenker (2005) for a comprehensive review. What has been lagging for quantile regression is reliable statistical inference tools. The inference problem is especially challenging for longitudinal data, as the asymptotic covariance of the quantile estimator involves not only the unknown error density function but also the unknown intra-subject correlation. Therefore, the conventional Wald-type methods based on the normal approximation and direct estimation of the covariance matrix is unstable in finite samples. Among limited literature on quantile regression for longitudinal data, Lipsitz et al. (1997) and Yin and Cai (2005) considered bootstrap methods by treating each subject (cluster) as the sampling unit, Chen et al. (2003) employed a resampling approach by reweighting the estimating equation, and Geraci and Bottai (2007) developed a likelihood-based approach assuming an asymmetric Laplace distribution. This paper employs an empirical likelihood (EL) approach to construct confidence regions of quantile coefficients.

Empirical likelihood (EL), introduced by Owen (1988), is a nonparametric inference method based on likelihood-ratio-type statistics through profiling a nonparametric likelihood. Chen and Hall (1993) discussed EL confidence intervals for the population quantiles (no covariates). For quantile regression models with cross sectional data, Whang (2006) considered a smoothed EL and discussed its higher-order properties, Otsu (2008) focused on the first-order approximation of a smoothed conditional EL approach, and several other authors considered EL for censored survival data; see Qin and Tsao (2003), Subramanian (2007), Zhao and Chen (2008) and Zhou et al. (2009).

* Corresponding author.
E-mail addresses: wang@stat.ncsu.edu (H.J. Wang), zhuzy@fudan.edu.cn (Z. Zhu).
We provide a new EL inference tool for quantile regression in longitudinal studies. For longitudinal data, the usual formulation of EL in the aforementioned literature is not applicable, as it would ignore the intra-subject correlation, and cause the empirical likelihood ratio statistic to lose its major attraction—the limiting chi-square distribution (You et al., 2006). To accommodate the intra-subject correlation, we employ the blocking technique, which was also used in Chen and Wong (2009) for quantiles of univariate weakly dependent data, and in You et al. (2006) and Xue and Zhu (2007) for semiparametric mean regression models.

In addition, to achieve higher-order accuracy, we propose a smoothed SEL procedure by replacing the quantile score function with a smoothed counterpart. The idea of smoothing quantile score function was also considered by Horowitz (1998), suitable conditions, the proposed SEL confidence regions have coverage errors of order $O(n^{-1})$, which can be reduced to $O(n^{-2})$ with a simple Bartlett correction. We want to emphasize that for longitudinal data, it is technically more challenging to establish the higher-order properties than those given in Whang (2006). The main reason is that observations within the same subject have a multivariate joint distribution, and this complicates the verification of the Cramér condition for the Edgeworth expansion.

Our proposal has several advantages over existing inference methods such as those based on the normal approximation or bootstrap. First, the shape and orientation of the EL confidence regions are determined automatically by the data. Second, the proposed method does not require estimating either the unknown error density function or the intra-subject correlation, for technical details are deferred to the Appendix.

The rest of the paper is organized as follows. We introduce the proposed EL and SEL estimators for longitudinal data in Section 2. Section 3 reports the main theoretical results, and describes the construction of confidence regions. We demonstrate the finite-sample performance of the proposed method through a simulation study in Section 4, and the analysis of an ophthalmology data in Section 5. Section 6 concludes the paper with some remarks. All technical details are deferred to the Appendix.

2. Empirical likelihood method

Consider the following marginal quantile regression model for longitudinal data:

$$y_{ij} = x_{ij}'\beta_0 + e_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i,$$

where $y_{ij}$ is the $j$th measurement of the $i$th subject, $x_{ij}$ is the observed $p$-dimensional design vector, $\beta_0$ is a $p$-vector of unknown parameters, $e_{ij}$ is the random error satisfying $P(e_{ij} < 0|x_{ij}) = \tau$ for any $i$ and $j$. Here $\tau \in (0, 1)$ is the quantile level of interest. The random errors are correlated within the same subject, but independent between subjects.

The standard quantile regression estimator $\hat{\beta}_0$ solves

$$\min_{\beta \in B} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho_i(y_{ij} - x_{ij}'\beta),$$

where $B$ is the parameter space and $\rho_i(u) = u(\tau - I(u < 0))$ is the quantile loss function. For independent data, Koenker and Bassett (1978) showed that $\hat{\beta}_0$ is $n^{1/2}$-consistent and asymptotically normal. Under the above model assumptions, $\hat{\beta}_0$ satisfies the following estimating equation:

$$E[\psi(y_i, x_i, \beta_0)|y_i, x_i, \beta_0] = 0,$$

where $\psi(y_i, x_i, \beta) = I(x_i'\beta - y_i > 0) - \tau$ is the quantile score function, and $I(\cdot)$ is the indicator function.

For longitudinal data, the usual EL formulation cannot be used to derive the desired Wilks’ theorem due to the correlation within subjects (You et al., 2006). To account for such intra-subject correlation, we employ a blocking technique by treating $(\psi(y_{ij}, x_{ij}, \beta))_{j=1, \ldots, n_i}$ as a whole unit in the development of EL. For the sake of convenience, we define $X_i = (x_{i_1}, \ldots, x_{in_i})^T$ as the $n_i \times p$ design matrix on the $i$th subject, $\psi_i(\beta) = (\psi(y_{i1}, x_{i1}, \beta), \ldots, \psi(y_{in_i}, x_{in_i}, \beta))^T$, and $Z_i(\beta) = X_i'^T\psi_i(\beta)$.

Let $p_1, \ldots, p_n$ be nonnegative numbers satisfying $\sum_{i=1}^{n} p_i = 1$. The block empirical log likelihood ratio for $\beta$ is defined as

$$l(\beta) = -2\max \left\{ \sum_{i=1}^{n} \log(p_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_i(\beta) = 0 \right\}.$$

(1)

Here we focus on the unbiased estimating equation assuming working independence. We can also incorporate the intra-subject dependence structure and heteroscedasticity in $Z_i(\beta)$ and form a weighted empirical likelihood; see Section 6 for more discussions.

From the standard Lagrange multiplier method, the optimal $p_i$ solving (1) for a given $\beta$ is

$$p_i(\beta) = n^{-1}(1 + \lambda(\beta)'Z_i(\beta))^{-1},$$

where $\lambda(\beta)$ is a $p$-dimensional Lagrange multiplier satisfying

$$n^{-1} \sum_{i=1}^{n} \frac{Z_i(\beta)}{1 + \lambda(\beta)'Z_i(\beta)} = 0.$$

(2)
Therefore, the empirical log likelihood ratio statistic can be written as

$$k(\beta) = 2 \sum_{i=1}^{n} \log \{1 + \lambda(\beta)^T Z_i(\beta)\}$$

with $\lambda(\beta)$ satisfying (2). Throughout our numerical studies, we solve (2) for $\lambda(\beta)$ by employing the modified Newton–Raphson algorithm of Chen et al. (2002). We define the maximum empirical likelihood (EL) estimator of $\beta_0$ as

$$\hat{\beta}_{EL} = \arg \min_{\beta \in B} k(\beta).$$

(3)

In empirical likelihood, higher-order refinements may be obtained by using Taylor series approximation, which requires smooth moment restrictions. However, the quantile score function $\psi(\cdot)$ is not differentiable at $\beta$ with $y = x^T \beta$. To achieve the higher-order accuracy, we consider a smooth empirical likelihood (SEL) approach by approximating $\psi(\cdot)$ by a smooth function $\psi_h(\cdot)$. Let $K(\cdot)$ denote a kernel function satisfying assumption A4 given in Section 3.1. Define $G(x) = \int_{-\infty}^{x} K(u) du$ and $G_h(x) = G(x/h)$, where $h$ is a positive bandwidth parameter. Then, we approximate $\psi(\cdot)$ by $\psi_h(y_i \mid x_i, \beta) = G_h(y_i - \beta - \gamma_i - \tau)$. Let $Z_h(\beta) = X_i^T \psi_h(\beta)$ and

$$l_h(\beta) = -2 \max \left\{ \sum_{i=1}^{n} \log(p_i), p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_h(\beta) = 0 \right\}.$$  

We define the maximum smooth empirical likelihood estimator of $\beta_0$ as

$$\hat{\beta}_{SEL} = \arg \min_{\beta \in B} l_h(\beta).$$

(4)

In Theorem 1 below, we show that under some regularity conditions, $\hat{\beta}_{EL}$ and $\hat{\beta}_{SEL}$ have the same asymptotic distribution as $\hat{\beta}_Q$, as $h$ goes to zero sufficiently fast when $n \to \infty$.

3. Main results

3.1. Estimation

We first derive the theoretical properties of the EL quantile estimators $\hat{\beta}_{EL}$ and $\hat{\beta}_{SEL}$. For the sake of convenience, we assume a balanced design with $n_1 = \ldots = n_m = m$. We denote $F(u_1, \ldots, u_m \mid x)$ as the joint distribution function of $e_i = (e_{i1}, \ldots, e_{im})^T$, and $f_j(u_j \mid x)$ as the marginal distribution function of $e_j$ conditional on $X_i = x$. We define $f(u_1, \ldots, u_m \mid x)$ as the joint density of $e_i$, and $f_j(u_j \mid x)$ as the marginal density of $e_j$ with respect to the Lebesgue measure. Furthermore, let $T(u \mid x) = \text{diag}(f_1(u_1 \mid x), \ldots, f_m(u_m \mid x))$, and $S = E(X_i^T T(0) X_i)$, and $\Sigma = E(X_i^T \psi(h) \psi(h)^T X_i)$, where $\psi(h) = (\psi(y_1 \mid x_1, \beta_0), \ldots, \psi(y_m \mid x_m, \beta_0))^T$. We let $r \geq 2$ be an integer and posit the following assumptions.

A1. Let $Y_i = (y_{i1}, \ldots, y_{im})^T$. The $\{Y_i, X_i\}_{i=1}^{n}$ are i.i.d. random vectors.

A2. The parameter vector $\beta_0$ is an interior point of the parameter space $B$, a compact subset of $R^p$.

A3. $X_i$ has a bounded support, and matrices $S$ and $\Sigma$ are nonsingular.

(a) $f(u_1, \ldots, u_m \mid x)$ has a continuous partial derivative with respect to $u_j$, $j = 1, \ldots, m$.

(b) For all $u_j$ in a neighborhood of 0 and almost every $x$, $f_j(u_j \mid x)$ exist, are bounded away from zero, and $r$ times continuously differentiable with respect to $u_j$, $j = 1, \ldots, m$.

A4. (a) $K(\cdot)$ is bounded and compactly supported on $[-1,1]$.

(b) For some constant $C_k \neq 0$, $K(\cdot)$ is an $r$ th-order kernel, i.e., $\int \! u K(u) du = 1$ if $j = 0$; 0 if $1 \leq j \leq r - 1$; $C_k$ if $j = r$.

(c) Let $G(u) = (G(u), G^2(u), \ldots, G^{r+1}(u))^T$ for some $L \geq 1$, where $G(u) = \int_{-u}^{u} K(v) dv$. For any $\theta \in R^{r+1}$ satisfying $\|\theta\| = 1$, there is a partition of $[-1,1]$, $-1 = a_0 < a_1 < \cdots < a_{L+1}$ such that $\theta^T G(u)$ is either strictly positive or strictly negative on $(a_{i-1}, a_i)$ for $i = 1, \ldots, L+1$.

A5. The positive bandwidth parameter $h$ satisfies (a) $nh^{2r} \to 0$ and (b) $nh / \log n \to \infty$ as $n \to \infty$.

The above assumptions are similar to those in Horowitz (1998) and Whang (2006), which were used to obtain the asymptotic refinement of bootstrap procedure based on the smoothed median estimator, and to study the asymptotic properties of the smoothed empirical likelihood method. In this paper, we use the above assumptions to establish the asymptotic results of the empirical likelihood method for longitudinal data. Longitudinal data are characterized by the dependence among observations within the same subject and independence among different subjects. The intra-subject dependence causes difficulty for studying the higher-order asymptotic properties of the smoothed empirical likelihood procedure.

We now present the asymptotic properties of the EL estimator $\hat{\beta}_{EL}$ and the smoothed EL estimator $\hat{\beta}_{SEL}$.

Theorem 1. (i) If assumptions A1–A3 hold, then $n^{1/2}(\hat{\beta}_{EL} - \beta_0) = o_p(1)$ and $n^{1/2}(\hat{\beta}_{EL} - \beta_0) \to N(0, \Omega)$, where $\Omega = S^{-1} \Sigma S^{-1}$.

(ii) If assumptions A1–A4(b) and A5(a) hold, then $n^{1/2}(\hat{\beta}_{SEL} - \beta_0) = o_p(1)$ and $n^{1/2}(\hat{\beta}_{SEL} - \beta_0) \to N(0, \Omega)$. 
The asymptotic covariance matrix $\Omega$ can be estimated by
\begin{equation}
\hat{\Omega} = \hat{S}^{-1} \hat{\Sigma} \hat{S}^{-1}, \quad \hat{S} = (nh)^{-1} \sum_{i=1}^{n} X_i^T \mathbf{K}_n(\hat{\beta}) X_i, \quad \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} X_i^T \psi(\hat{\beta}) \psi(\hat{\beta})^T X_i,
\end{equation}
where $\mathbf{K}_n(\hat{\beta}) = \text{diag}(K(y_i - x_i^T \hat{\beta})/h), \ldots, K(y_m - x_m^T \hat{\beta})/h)$, $\hat{\beta}$ is any consistent estimator of $\beta_0$. Under the same assumptions of Theorem 1, it is easy to show that $\hat{\Omega}$ is consistent with $\Omega$.

### 3.2. Hypothesis test and confidence region

We next derive a nonparametric analog of the Wilks’ Theorem.

**Theorem 2.** (i) If assumptions A1–A3 hold, then $l(\beta_0) \rightarrow \chi^2(p)$. (ii) If assumptions A1–A4(b) and A5(a) hold, then $l_0(\beta_0) \rightarrow \chi^2(p)$.

As an analogy to parametric likelihoods, Theorem 2 allows us to use the test statistics $l(\beta_0)$ and $l_0(\beta_0)$ for testing or obtaining confidence regions for $\beta_0$. Specifically, we define
\begin{equation}
l_0 = \{\beta : l(\beta) \leq \chi^2_{1-\alpha}(p)\}, \quad l_0 = \{\beta : l_0(\beta) \leq \chi^2_{1-\alpha}(p)\},
\end{equation}
as the EL and SEL confidence regions for $\beta_0$, respectively, where $\chi^2_{1-\alpha}(p)$ is the $(1-\alpha)$th quantile of $\chi^2(p)$.

When interest lies on a subset of the parameters, profile empirical log likelihood ratio test statistic can be used to test or obtain a confidence region for the subset in the following way. Let $\beta_0 = (\beta_0^{(1)T}, \beta_0^{(2)T})^T$, where $\beta_0^{(1)}$ and $\beta_0^{(2)}$ are $q \times 1$ and $(p-q) \times 1$ vectors, respectively. Suppose we are interested in testing $H_0 : \beta_0^{(1)} = \tilde{\zeta}$, where $\tilde{\zeta}$ is some known $q \times 1$ vector. The profile log likelihood ratio test statistic is defined as
\begin{equation}
l_{h2}(\zeta) = l_{h}(\tilde{\zeta}, \beta^{(2)}),
\end{equation}
where $h = 0$ corresponds to the EL, and $h = h > 0$ corresponds to the SEL. $\beta^{(2)}$ minimizes $l_h(\tilde{\zeta}, \beta^{(2)})$ with respect to $\beta^{(2)}$. $(\beta^{(1)}, \beta^{(2)})$ are either the EL or SEL estimators defined in Section 3.1. Following the similar arguments as in the proof of Theorem 2, we can show the following corollary.

**Corollary 1.** Under the assumptions of Theorem 2 and $H_0 : \beta_0^{(1)} = \tilde{\zeta}$, we have $l_{h2}(\tilde{\zeta}) \rightarrow \chi^2(q)$, as $n \rightarrow \infty$, where $h = 0$ or $h$.

An approximate $(1-\alpha)$ confidence region for $\beta_0^{(1)}$ thus can be obtained by
\begin{equation}
l_h = \{\tilde{\zeta} : \tilde{\zeta} \in \mathbb{R}^q, l_{h2}(\tilde{\zeta}) \leq \chi^2_{1-\alpha}(q)\}, \quad h = 0 \text{ or } h = h > 0.
\end{equation}

### 3.3. Bartlett correction of the smoothed empirical likelihood

The following theorem indicates that the coverage error of the confidence region $l_h$ obtained by using the chi-square approximation to the distribution of $l_h(\beta_0)$ is of order $O(n^{-1})$.

**Theorem 3.** Suppose assumptions A1–A5 hold and $\sup_{n} nh^r < \infty$, then $P(l_0 \in l_h) = 1-\alpha + O(n^{-1})$.

It can be proven that if $h > 0$ and $nh^r \rightarrow 0$, then
\begin{equation}
E(l_h(\beta_0)) = p(1 + n^{-1}b) + O(n^{-1}),
\end{equation}
where $b$ is the Bartlett factor to be explained shortly. Note that the expected value of $l_h(\beta_0)$ differs from that of its approximate chi-square distribution by $n^{-1}pb$. By correcting this bias, it is reasonable to expect the chi-square approximation to be more accurate. This is the idea of the Bartlett correction.

To study the higher-order property of the SEL confidence regions, we employ an Edgeworth-type expansion for the distribution of the empirical log likelihood ratio statistic. To justify the Edgeworth expansion, we need to verify a modified version of the Cramér condition; see Lemma 2. This verification is technically challenging for longitudinal data, as in such data the observations within the same subject have a multivariate joint distribution, which makes the Cramér condition take a more complicated form than for the univariate distribution.

For easy demonstration, we standardize $\lambda = \lambda(\beta_0)$ and $Z_{hi} = Z_{hi}(\beta_0)$ to
\begin{equation}
t = V_{n}^{1/2} \lambda \quad \text{and} \quad W_i = V_{n}^{-1/2} Z_{hi}, \quad i = 1, \ldots, n,
\end{equation}
respectively, where $V_n = E(Z_{hi}Z_{hj})$. Then, the empirical log likelihood ratio statistic $l_h(\beta_0)$ can be rewritten as
\begin{equation}
l_h(\beta_0) = 2 \sum_{i=1}^{n} \log(1 + t^T W_i),
\end{equation}
where \( t \) satisfies
\[
\frac{1}{n} \sum_{i=1}^{n} W_i = 0.
\]

Let \( W_i \) denote the \( j \)-th component of \( W_i \). We adopt the following \( z - A \) system of notations used in McCullagh (1987), DiCiccio et al. (1991), and Chen and Cui (2006):

\[
\gamma^{h-j} = E(W_{1i} \cdots W_{ni}), \quad A^{h-j} = n^{-1} \sum_{i=1}^{n} W_{1i} \cdots W_{ni} \text{, and} \quad A^{h-j} = \gamma^{h-j} - \gamma^{h-j}.
\]

Similar to Whang (2006), we can show that the Bartlett factor
\[
b = p^{-1}(x^{ikk}/2 - \check{x}^{ikk}x^{ikk}/3).
\]

By (7), we obtain a corrected \( (1-\alpha) \) confidence region
\[
I^h_b = \{ \beta : l_b(\beta) \leq \chi^2_{1-\alpha}((1+n^{-1})b) \}.
\]

Similarly, for testing \( H_0 : \beta_0 = \check{\beta}_0 \) in Section 3.2, a corrected level \( z \) test will reject \( H_0 \) when \( l_0(\beta_0) > \chi^2_{1-\alpha}(q(1+n^{-1})b) \).

In practice, \( b \) is unknown and has to be estimated. Let \( \hat{\beta} \) denote any \( n^{1/2} \)-consistent estimator of \( \beta_0 \) such as \( \hat{\beta}_E \) and \( \hat{\beta}_Q \). We can first obtain a consistent estimator of \( V_n, \hat{V}_n(\hat{\beta}) = n^{-1} \sum_{i=1}^{n} Z_n(\hat{\beta})Z_n(\hat{\beta})^T \). We then define the estimated Bartlett factor as
\[
b = p^{-1}(\check{x}^{ikk}/2 - \check{x}^{ikk}x^{ikk}/3),
\]

where
\[
\check{x}^{ikk} = n^{-1} \sum_{i=1}^{n} (Z_n(\hat{\beta})_n(\hat{\beta})^T)^2 \quad \text{and} \quad \check{x}^{ikk}x^{ikk} = n^{-1} \sum_{i=1}^{n} |Z_n(\hat{\beta})_n(\hat{\beta})^T|^2.
\]

The following theorem states that with the Bartlett correction, the accuracy of the SEL confidence region can be improved to the order of \( O(n^{-2}) \).

**Theorem 4.** Suppose assumptions A1–A5 hold and \( \sup_n n^3h^{2r} < \infty \). Then as \( n \to \infty \), we have
(a) \( P(\beta_0 \in I^h_b) = 1-\alpha + O(n^{-2}) \); (b) \( P(\beta_0^B \in I^b_0) = 1-\alpha + O(n^{-2}) \).

**4. Simulation study**

We carry out a simulation study to investigate the proposed EL approach. We consider two different models.

**Model 1 (homoscedastic):**
\[
y_{ij} = \beta_1 + x_{ij}\beta_2 + e_{ij}(\tau), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m,
\]
where \( e_{11}, \ldots, e_{nm} \sim N(0, V) \). \( V \) has an exchangeable correlation structure with diagonal entries 1 and off-diagonal entries 0.7, \( e_{ij}(\tau) = e_{ij} - \Phi^{-1}(\tau) \) with \( \Phi \) as the cumulative distribution function of \( N(0,1) \). Here \( \Phi^{-1}(\tau) \) is subtracted from \( e_{ij} \) so that the \( \tau \)-th quantile of \( e_{ij}(\tau) \) is zero.

**Model 2 (heteroscedastic):**
\[
y_{ij} = \beta_1 + x_{ij}\beta_2 + 0.25(1 + |x_{ij}|)e_{ij}(\tau), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m,
\]
where \( e_{11}, \ldots, e_{nm} \sim N(0, V) \). \( V \) has an AR(1) correlation structure, i.e., \( \text{Corr}(e_{ij}, e_{ik}) = 0.7^{|i-j|} \), and \( e_{ij}(\tau) = e_{ij} - \Phi^{-1}(\tau) \). In both Models 1 and 2, we let \( m = 10, \beta_0 = \beta_1 = 1, \) and \( x_{ij} \sim N(0.3j, 0.5^2) \). In this simulation study, we focus on \( \tau = 0.5 \) and 0.7, and we choose \( n = 30 \) and 50 representing small and moderate sample sizes. The number of simulation replications is set to 2000.

We consider confidence intervals for \( \beta_1 \) and \( \beta_2 \), and confidence regions for the parameter vector \( (\beta_1, \beta_2)^T \). Throughout, we smooth the EL using the second-order Bartlett kernel (i.e., \( r = 2 \)).

\[
K(u) = \frac{3}{4\sqrt{5}}(1-u^2/5)I(|u| \leq \sqrt{5}).
\]

Let \( EL \) denote the unsmoothed empirical likelihood confidence interval (region), \( SEL \) be the smoothed empirical likelihood confidence interval (region), and \( SEL^b \) be the Bartlett corrected counterpart of \( SEL \). As comparisons, we also include another two methods \( Wald \) and \( Boot \), for which the large sample confidence intervals (regions) are obtained by using the asymptotic normality of the standard (unsmoothed) estimator \( \beta_0 \). In \( Wald \), the asymptotic covariance of \( \hat{\beta}_Q \), \( \Omega \), is estimated directly by \( \Omega \) defined in (5). In \( Boot \), \( \Omega \) is estimated by the sample covariance matrix of 2000 bootstrap estimates \( \beta_0 \), from the 2000 bootstrap samples, each of which is obtained by resampling the \((Y_i, X_i)\) pairs with replacement treating each subject as a whole unit. In \( Wald \) and \( Boot \), the standard estimator \( \beta_0 \) is obtained by using the “rq” function in R package quantreg.

The methods \( SEL, SEL^b \) and \( Wald \) all rely on the smoothing parameter \( h \). To study the sensitivity of three methods to the smoothing parameter in finite samples, we consider \( h = n^{-\delta} \), and \( \delta \in \{0.1, 0.2, \ldots, 1.0\} \). Fig. 1 depicts the coverage probabilities...
Fig. 1. Coverage probabilities of 95% confidence intervals for $\beta_2$ obtained from SEL (curve with open circles), SEL$^b$ (curve with solid dots), and Wald (curve with triangles) with bandwidth $h = n^{-1}$. The two dotted horizontal lines are the Bonferroni corrected 95% confidence band for the coverage probability.

Table 1
Estimated coverage probabilities (CP) of confidence intervals (regions) for $\beta_1$, $\beta_2$, and $(\beta_1, \beta_2)^T$, and the mean lengths (ML) of confidence intervals from different methods in Model 1.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\tau$</th>
<th>Method</th>
<th>$\beta_1$ CP</th>
<th>$\beta_1$ ML</th>
<th>$\beta_2$ CP</th>
<th>$\beta_2$ ML</th>
<th>$(\beta_1, \beta_2)$ CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>EL</td>
<td>0.936</td>
<td>0.613</td>
<td>0.931</td>
<td>0.213</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL</td>
<td>0.936</td>
<td>0.631</td>
<td>0.942</td>
<td>0.203</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL$^b$</td>
<td>0.940</td>
<td>0.673</td>
<td>0.948</td>
<td>0.208</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.929</td>
<td>0.663</td>
<td>0.945</td>
<td>0.219</td>
<td>0.902</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boot</td>
<td>0.930</td>
<td>0.661</td>
<td>0.964</td>
<td>0.231</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>EL</td>
<td>0.931</td>
<td>0.648</td>
<td>0.929</td>
<td>0.224</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL</td>
<td>0.930</td>
<td>0.644</td>
<td>0.925</td>
<td>0.212</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL$^b$</td>
<td>0.937</td>
<td>0.694</td>
<td>0.933</td>
<td>0.218</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.926</td>
<td>0.683</td>
<td>0.933</td>
<td>0.231</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boot</td>
<td>0.927</td>
<td>0.685</td>
<td>0.963</td>
<td>0.246</td>
<td>0.913</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>EL</td>
<td>0.952</td>
<td>0.498</td>
<td>0.942</td>
<td>0.165</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL</td>
<td>0.955</td>
<td>0.498</td>
<td>0.947</td>
<td>0.161</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL$^b$</td>
<td>0.956</td>
<td>0.519</td>
<td>0.952</td>
<td>0.163</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.949</td>
<td>0.515</td>
<td>0.945</td>
<td>0.169</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boot</td>
<td>0.949</td>
<td>0.515</td>
<td>0.960</td>
<td>0.177</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>EL</td>
<td>0.950</td>
<td>0.513</td>
<td>0.932</td>
<td>0.174</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL</td>
<td>0.948</td>
<td>0.510</td>
<td>0.940</td>
<td>0.167</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEL$^b$</td>
<td>0.950</td>
<td>0.540</td>
<td>0.944</td>
<td>0.170</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.940</td>
<td>0.531</td>
<td>0.941</td>
<td>0.177</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boot</td>
<td>0.943</td>
<td>0.532</td>
<td>0.958</td>
<td>0.186</td>
<td>0.929</td>
</tr>
</tbody>
</table>
of 95% confidence intervals for $\beta_2$ obtained from $SEL$, $SEL^b$, and $Wald$ against $\delta$ at $\tau = 0.5$. The curves with open circles, solid dots, and triangles correspond to $SEL$, $SEL^b$, and $Wald$, respectively. The dashed lines are the 95% confidence band for the confidence probability after Bonferroni correction accounting for the 10 $\delta$ values.

Fig. 1 shows that the $Wald$ method is very sensitive to the smoothing parameter. The coverage probabilities of $Wald$ lie generally outside the confidence band except for $\delta = 0.4, 0.5$ and 0.6. The $SEL$ and $SEL^b$ are relatively robust to the bandwidth, and both give decent coverage probabilities for $h \in [n^{-0.8}, n^{-0.9}]$. The Bartlett-corrected method $SEL^b$ has a higher coverage accuracy than $SEL$, suggesting an improvement offered by the Bartlett correction.

Tables 1 and 2 summarize the estimated coverage probabilities (CP) of confidence intervals (regions) for $\beta_1$, $\beta_2$ and $(\beta_1, \beta_2)^\top$, and the mean lengths (ML) of confidence intervals from different methods. Recall from Theorems 3 and 4 that when $r=2$, the uncorrected and Bartlett-corrected $SEL$ regions have coverage errors $O(n^{-1})$ and $O(n^{-2})$ if $h$ has order smaller than $n^{-1/2}$ and $n^{-3/4}$, respectively. We report the results of $SEL$ and $SEL^b$ with $h = n^{-0.8}$, and these of $Wald$ with $h = n^{-0.5}$, at which $Wald$ performs relatively better (see Fig. 1). Our findings are as follows. (1) The $EL$, $SEL$ and $SEL^b$ all give coverage probabilities close to the nominal level of 95% in all the situations considered, but $SEL^b$ has higher coverage accuracy for small samples with $n=30$. (2) The $Wald$ has lower coverage for $(\beta_1, \beta_2)^\top$, especially in Model 1. (3) The $Boot$ gives reasonable coverage probabilities at $n=50$, but it has low coverage probabilities at $n=30$ for $\beta_1$ (CP is around 93%), and for $(\beta_1, \beta_2)^\top$ (CP is around 91%). (4) When coverage probabilities are comparable, the $EL$-based confidence intervals are generally shorter than those of $Boot$ and $Wald$.

5. Application to an ophthalmology data

We now demonstrate the proposed methodology by analyzing an ophthalmology data set (Song and Tan, 2000). Before retinal repair surgeries, intraocular gas was injected into the eyes of 31 patients to provide internal tamponade of retinal breaks. Within three months after surgery, patients were followed up for 3–8 times, and the volume of gas left in the eye was measured and recorded as the percentage of the initial gas volume in that eye. Song and Tan (2000) analyzed the conditional means of the longitudinal remaining gas percentages. Instead, we focus on studying how the conditional quantiles of gas decays with time.

Let $y_{ij}$ be the gas volume left in the eye of patient $i$ at day $t_{ij}$. We define the logit-transformed response

$$\tilde{y}_{ij} = \log\left(\frac{y_{ij} + 0.05}{1 - y_{ij} + 0.05}\right),$$

where the constant 0.05 is added to avoid zero denominators. Let $x_i$ be the centered gas concentration of the $i$th subject, so that $x_i = -1, 0, 1$ corresponding to gas concentration levels of 15, 20 and 25, respectively. We consider the following quantile regression model

$$\tilde{y}_{ij} = \beta_0(t) + \beta_1(t)\log(t_{ij}) + \beta_2(t)\log^2(t_{ij}) + \beta_3(t)x_i + \epsilon_{ij}(t),$$

where the $r$th conditional quantile of $\epsilon_{ij}(t)$ given the other covariates is zero.
Compared to the mean regression method in Song and Tan (2000), quantile regression enjoys three appealing features for analyzing this data set. First, quantile regression does not assume any error distribution form, thus it is able to accommodate more flexible error distributions. Second, it is known that quantiles, unlike mean, are equivariant to monotone transformations. Therefore, the \( t \)th conditional quantile of \( y_{ij} \) can be obtained by transforming that of \( \tilde{y}_{ij} \) back to the original percentage scale, while such re-transformation for the mean often leads to biased results (Welsh and Zhou, 2006). In addition, analysis at different quantiles can provide a profile of covariate effects, and capture the effects overlooked by ordinary mean methods.

Fig. 2 demonstrates the profile of covariate effects at different quantiles with \( \tau \) ranging from 0.05 to 0.95. The solid line with open circles in each panel represents the point estimates, and the shaded area depicts a 95% pointwise confidence band obtained from the Bartlett-corrected method \( SEL_b \) with \( h = n^{-0.8} \). The horizontal dashed line represents the mean effects, and the dotted lines are the corresponding 95% confidence intervals from Song and Tan (2000). Consistent with the mean, \( SEL^b \)

![Graph](image)

**Fig. 2.** The point estimates (open circles) and 95% pointwise confidence band (shaded area) of the quantile coefficients. The horizontal dashed line and two dotted lines represent the point estimate, the 95% confidence interval for the mean coefficient, respectively.

![Graph](image)

**Fig. 3.** Scatter plot of data with the estimated conditional quantile functionals at \( \tau = 0.1, 0.5 \) and 0.9.
Table 3
Estimation and confidence intervals of quantile coefficients in the ophthalmology study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>95% confidence interval</th>
<th>Wald</th>
<th>Boot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SELb</td>
<td></td>
</tr>
<tr>
<td>( \tau = 0.1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.488</td>
<td>(0.206, 0.799)</td>
<td>0.016, 1.104</td>
<td>( -0.011, 1.352 )</td>
</tr>
<tr>
<td>Time (log( t ))</td>
<td>-0.251</td>
<td>(-0.640, 0.090)</td>
<td>-1.448, 0.857</td>
<td>-0.707, 0.764</td>
</tr>
<tr>
<td>Time(^2) (log( t ))</td>
<td>-0.128</td>
<td>(-0.192, -0.034)</td>
<td>-0.432, 0.188</td>
<td>-0.406, -0.015</td>
</tr>
<tr>
<td>Gas (x)</td>
<td>0.623</td>
<td>(0.201, 0.918)</td>
<td>0.021, 1.306</td>
<td>( -0.033, 1.253 )</td>
</tr>
<tr>
<td>( \tau = 0.5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.049</td>
<td>(1.568, 2.504)</td>
<td>1.087, 2.605</td>
<td>1.358, 2.681</td>
</tr>
<tr>
<td>Time (log( t ))</td>
<td>0.246</td>
<td>(-0.129, 0.666)</td>
<td>-0.158, 0.89</td>
<td>-0.202, 0.75</td>
</tr>
<tr>
<td>Time(^2) (log( t ))</td>
<td>-0.316</td>
<td>(-0.386, -0.227)</td>
<td>-0.425, -0.242</td>
<td>-0.415, -0.214</td>
</tr>
<tr>
<td>Gas (x)</td>
<td>0.413</td>
<td>(0.205, 0.725)</td>
<td>0.068, 0.765</td>
<td>0.084, 0.742</td>
</tr>
<tr>
<td>( \tau = 0.9 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.147</td>
<td>(3.048, 3.345)</td>
<td>2.797, 3.292</td>
<td>2.853, 3.445</td>
</tr>
<tr>
<td>Time (log( t ))</td>
<td>0.462</td>
<td>(0.462, 0.561)</td>
<td>0.131, 0.726</td>
<td>0.116, 0.85</td>
</tr>
<tr>
<td>Time(^2) (log( t ))</td>
<td>-0.308</td>
<td>(-0.339, -0.302)</td>
<td>-0.379, -0.189</td>
<td>-0.474, -0.199</td>
</tr>
<tr>
<td>Gas (x)</td>
<td>-0.094</td>
<td>(-0.441, 0.264)</td>
<td>-0.367, 0.367</td>
<td>-0.447, 0.725</td>
</tr>
</tbody>
</table>

shows that the log-transformed time has a significant quadratic effect at all quantile levels. This quadratic time effect tends to dominate over the linear effect, which is significant only at two tails. Song and Tan (2000) gave a \( p \)-value of 0.087 for testing the gas concentration effect on the mean gas volume. Our method, however, shows that gas concentration effect is significant (at the 5% level) for \( \tau \leq 0.6 \), that is, for the group of patients receiving less gas injection.

The fitted conditional quantiles of remaining gas (in percentage) are shown in Fig. 3 at \( \tau = 0.1, 0.5 \) and 0.9. In each panel, the dotted, solid and dashed curves correspond to the gas concentration levels 15, 20 and 25, respectively. The gas amount drops rapidly during the first 20 days at \( \tau = 0.1 \), and the first 40 days at \( \tau = 0.5 \), and it decreases slowly afterwards. As suggested by Fig. 2, gas concentration has an obvious positive effect at \( \tau = 0.1 \) and 0.5, suggesting that the higher gas concentration, the longer time it takes for the gas to drop to a certain level. For people receiving higher gas injection during surgery, i.e., those at the upper quantiles, the gas (in percentage) drops more slowly, and difference among gas concentration levels is ignorable.

We summarize the coefficient estimation and 95% confidence intervals obtained from \( SEL^b \), Wald, and the bootstrap method (Boot) based on 2000 bootstrap replications at \( \tau = 0.1, 0.5 \) and 0.9 in Table 3. As suggested by the first-order equivalency stated in Theorem 1, the EL estimator \( \hat{\beta}_{EL} \) and the standard quantile estimator \( \hat{\beta}_{Q} \) led to similar point estimates for this data set. Therefore, we only report the estimates from \( \hat{\beta}_{EL} \) in the second column of Table 3. The three inference methods agree in terms of significance of effects at \( \tau = 0.5 \) and 0.9. But at \( \tau = 0.1 \), Wald fails to identify the significant quadratic time effect, while Boot does not detect the intercept and gas concentration effects. In addition, the confidence intervals from \( SEL^b \) are universally narrower than those from Wald and Boot, suggesting higher testing efficiency.

6. Concluding remarks

We developed two block empirical likelihood-based inference procedures for quantile regression models with longitudinal data. The proposed methods do not require estimating the limiting covariance matrix, which involves both the unknown error density function and intra-subject correlation. As noted by Owen (1988) and Chen and Hall (1993) in the context of quantile regression, straight construction of empirical likelihood offers little extra over existing inference techniques such as bootstrap. The theoretical advantage of empirical likelihood lies in its potential higher-order efficiency. We demonstrated that by appropriately smoothing the quantile score function, the block empirical likelihood is Bartlett correctable. The establishment of the higher-order properties is particularly challenging, as the longitudinal data feature complicates the verification of the Cramér condition for the Edgeworth expansion. We provided a simple Bartlett correction that improves the coverage accuracy from \( n^{-1} \) to \( n^{-2} \). The simulation study confirmed that the Bartlett-corrected EL confidence regions had better coverage than those from the normal-approximation-based and bootstrap methods, and the EL counterparts without Bartlett correction. In addition, the finite-sample performance of the smoothed EL method was very robust to the choices of smoothing parameters.

Besides higher-order refinement, empirical likelihood has another potential advantage when the number of moment constraints is greater than the number of parameters; see Qin and Lawless (1994). Further investigation is needed to study how to automatically combine information about a distribution through empirical likelihood in quantile regression to improve efficiency. In this paper, our proposed method is based on the estimating equation assuming working independence. Efficiency might be gained by formulating a weighted empirical likelihood inference procedure by incorporating both the intra-subject dependence structure and heteroscedasticity in the estimating equation, as done in Jung (1996) under a
different setup. Our preliminary studies suggested that in quantile regression with finite samples, more efficiency can be gained by accounting for the heteroscedasticity than for the intra-subject correlation, unless the correlation is extremely high. We leave theoretical and empirical investigation of this matter to the future research.

Acknowledgements

The authors would like to thank two anonymous reviewers for constructive comments and helpful suggestions. The research is supported by the National Science Foundation Awards DMS-0706963, DMS-1007420, and National Natural Science Foundation of China Grant 10931002, 1091120386.

Appendix

Lemma 1. Under assumptions A1–A4(b) and A5(a), we have, as \( n \to \infty \),

\[
(a) \ E\mathbb{Z}_n(\beta_0) = (-h)^\gamma(\mathbb{Z})^{-1} \mathcal{E}[X_i f^{(r-1)}(0|X_i)] + o(h^{\gamma});
\]

\[
(b) \ E\mathbb{Z}_n h(\beta_0) Z^T_n(\beta_0) = \mathbf{\Sigma} + o(1);
\]

\[
(c) \ E\mathbb{Z}_n h(\beta_0) / \mathbb{\Sigma} = S + o(1).
\]

where \( f^{(r-1)}(0|X_i), \ldots, f^{(r-1)}(0|X_i) \).

Proof. The results can be obtained by using a standard technique for deriving the biases and variances of kernel density estimator. \( \square \)

Proof of Theorem 1(i). First, applying Lemma 4.1 of He and Shao (1996) gives

\[
n^{-1} \sum_{i=1}^{n} Z_i(\beta) = n^{-1} \sum_{i=1}^{n} Z_i(\beta_0) + n^{-1} \sum_{i=1}^{n} X_i f(0|X_i) X_i(\beta_0) + o_p(n^{-\delta}) = D_n + S(\beta_0) + o_p(n^{-\delta})
\]

uniformly on \( \Theta_n = \{ \beta : \| \beta - \beta_0 \| \leq n^{-\delta} \} \), where \( \delta > 0 \) and \( D_n = n^{-1} \sum_{i=1}^{n} Z_i(\beta_0) \).

By (10) and similar arguments of the proof of Lemma 1 in Qin and Lawless (1994), we have

\[
\hat{\beta}(\beta) = \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta) Z_i(\beta)^T \right\}^{-1} n^{-1} \sum_{i=1}^{n} Z_i(\beta) + o_p(n^{-\delta}) = \mathbf{\Sigma}^{-1} n^{-1} \sum_{i=1}^{n} Z_i(\beta) + o_p(n^{-\delta}) = o_p(n^{-1/2} + n^{-\delta})
\]

uniformly on \( \beta \in \Theta_n \). By assumption A3, (10) and (11) and the Taylor expansion,

\[
l(\beta) = 2 \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta) Z_i(\beta)^T - 1/2 \sum_{i=1}^{n} (\hat{\beta}(\beta) Z_i(\beta)^T)^2 + o_p(1 + n^{-1-2\delta}) \right\}
\]

\[
= n \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta) \right\}^T \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta) Z_i(\beta)^T \right\}^{-1} \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta) \right\} + o_p(1 + n^{1/2-\delta} + n^{-2\delta})
\]

uniformly on \( \beta \in \Theta_n \). Using similar arguments as above, we obtain

\[
l(\beta) = n \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta_0) \right\}^T \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta_0) Z_i(\beta_0)^T \right\}^{-1} \left\{ n^{-1} \sum_{i=1}^{n} Z_i(\beta_0) \right\} + o_p(1).
\]

By the central limit theorem, we have

\[
n^{1/2} D_n \to \mathcal{N}(0, \mathbf{\Sigma}).
\]

Similar to the proof of (28)–(30) in Otsu (2008), we have \( \hat{\beta}_E - \beta_0 = o_p(n^{-1/2}) \). By (12) and (13),

\[
l(\hat{\beta}_E) - l(\beta_0) = 2n D_n^T S(\hat{\beta}_E - \beta_0) + n(\hat{\beta}_E - \beta_0)^T S^{-1} S(\hat{\beta}_E - \beta_0) + o_p(1).
\]

Similarly,

\[
l(\hat{\beta}) - l(\beta_0) = 2n D_n^T S(\hat{\beta} - \beta_0) + n(\hat{\beta} - \beta_0)^T S^{-1} S(\hat{\beta} - \beta_0) + o_p(1).
\]

By the above two equations,

\[
l(\hat{\beta}) - l(\beta) = n(\hat{\beta} - \beta) S^{-1} S(\hat{\beta} - \beta) + o_p(1) \leq 0.
\]

Therefore,

\[
n^{1/2}(\hat{\beta} - \beta_0) = n^{1/2}(\hat{\beta} - \beta_0) + o_p(1) = -n^{1/2}(S^{-1} S)^{-1} S D_n + o_p(1) \to \mathcal{N}(0, S^{-1} S).
\]

\( \square \)
Proof of Theorem 1(ii). By assumptions A2 and A3 and the similar arguments used in the proof of Lemma 1 of Qin and Lawless (1994), we can show that with probability 1 as \( n \to \infty \), there exists a vector \( \hat{\beta}_{\text{SEL}} \in \text{int}(B) \) with \( \| \hat{\beta}_{\text{SEL}} - \beta_0 \| \leq Cn^{-1/3 + \varepsilon} \), for any \( c > 0, \varepsilon > 0 \), such that \( l_n(\beta) \) attains its minimum value at \( \hat{\beta}_{\text{SEL}} \). Here \( \hat{\beta}_{\text{SEL}} \) satisfies \( \lambda(\hat{\beta}_{\text{SEL}}) = 0 \) and \( Q_n(\hat{\beta}_{\text{SEL}}) = 0 \), where \( Q_n(\beta) = n^{-1} \sum_{i=1}^{n} Z_{hi}(\beta) \).

Applying Taylor expansion to \( Q_n(\hat{\beta}_{\text{SEL}}) \) around \( \beta_0 \), we obtain

\[
\hat{\beta}_{\text{SEL}} - \beta_0 = -\left( \frac{\partial^2 Q_n(\beta_0)}{\partial \beta^2} \right)^{-1} Q_n(\beta_0) + o_p(n^{-1/2}).
\]

By the law of large numbers,

\[
\frac{\partial^2 Q_n(\beta_0)}{\partial \beta^2} \to 0 \text{ in probability.}
\]

Using assumptions A3, A4(a), (b) and A5(a), similar to Lemma 3(k) of Horowitz (1998), we obtain

\[
n^{1/2} Q_n(\beta_0) = -n^{-1/2} \sum_{i=1}^{n} X_i^T (l(e_i \leq 0) - 1) + o_p(1),
\]

(16)

where \( l(e_i \leq 0) = (l(e_{i1} \leq 0), \ldots, l(e_{im} \leq 0))^T, 1 = (1, \ldots, 1)^T \). The proof of Theorem 1 thus completes by noting that (16) is the Bahadur representation of the usual quantile regression estimator \( \beta_{Q} \). □

Proof of Theorem 2. First it follows directly from (13) and (14) that \( 2l(\beta_0) \to Z_{\rho}^2 \). Next we will prove Theorem 2(ii). Let \( \hat{\lambda} = \lambda(\beta_0) \) denote the solution of the equation

\[
n^{-1} \sum_{i=1}^{n} \frac{Z_{hi}(\beta_0)}{1 + Z_{hi}(\beta_0)} = 0.
\]

By Lemma 1 and similar arguments to those used in Owen (1990), we have

\[
\hat{\lambda} = o_p(n^{-1/2} + h').
\]

Furthermore, by assumption A3, we get

\[
\hat{\lambda} = \left( n^{-1} \sum_{i=1}^{n} Z_{hi} Z_{hi}^T \right)^{-1} \left( n^{-1} \sum_{i=1}^{n} Z_{hi} \right) + o_p(n^{-1/2} + h').
\]

Note that

\[
l_n(\beta_0) = 2 \sum_{i=1}^{n} \log(1 + \hat{\lambda}^T Z_{hi}) = 2 \sum_{i=1}^{n} \hat{\lambda}^T Z_{hi} - \sum_{i=1}^{n} (\hat{\lambda}^T Z_{hi})^2 + o_p(1)
\]

\[
= \left( n^{-1/2} \sum_{i=1}^{n} Z_{hi} \right)^T \left( n^{-1} \sum_{i=1}^{n} Z_{hi} Z_{hi}^T \right)^{-1} \left( n^{-1} \sum_{i=1}^{n} Z_{hi} \right) + o_p(1)
\]

\[
= \left( n^{-1/2} \sum_{i=1}^{n} (Z_{hi} - EZ_{hi}) \right)^T \left( n^{-1} \sum_{i=1}^{n} Z_{hi} Z_{hi}^T \right)^{-1} \left( n^{-1/2} \sum_{i=1}^{n} (Z_{hi} - EZ_{hi}) \right) + o_p(1).
\]

By a Taylor expansion, Lemma 1(a), and assumption A5(a), \( n^{1/2} EZ_{hi} \to 0 \), and \( (n^{-1} \sum_{i=1}^{n} Z_{hi} Z_{hi}^T)^{-1/2} (n^{-1/2} \sum_{i=1}^{n} (Z_{hi} - EZ_{hi}) - 0) \to \mathcal{N}(0, I_p) \). Therefore, \( l_n(\beta_0) \) has an asymptotic chi-square distribution with \( p \) degrees of freedom. □

The proof of Theorem 3 is based on the following lemma, which states a modified version of Cramér’s condition. This condition is very important for the Edgeworth expansion used later.

For any \( L > 1 \), let

\[
\overline{Q} = (A^1, \ldots, A^p, A^{11}, \ldots, A^{pp}, A^{11}, \ldots, A^{pp}, p)^T = n^{-1} \sum_{i=1}^{n} Q_i
\]

denote a vector of all distinct first \( L + 1 \) order multivariate centered moments of \( W_i = V_n^{-1/2} Z_{hi} \).

Lemma 2. Let \( t \in \mathbb{R}^{l(m)} \) be a vector and \( l(t, h) = E(\exp(it Q)) \), where \( Q = (Q_i) \) and \( i = \sqrt{-1} \). Under assumptions A1–A5, there exists some \( C(\varepsilon) > 0 \) for each \( \varepsilon > 0 \) such that

\[
\sup_{|t| < \varepsilon} \left| l(t, h) \right| < 1 - C(\varepsilon) h \quad \text{for all sufficiently small } h.
\]

(17)

Proof. The proof of Lemma 2 is analogous to those of Horowitz (1998) and Whang (2006). Let \( g_{\varepsilon}(X) \) be a vector of the products of \( X \) corresponding to the \( \nu \)th order polynomial \( |G(-e_i/h)|^\nu \) in the expansion of \( t^Q \). Let \( q_{\varepsilon}(t) \) denote the
corresponding subvector of $t \in \mathbb{R}^{\dim(Q)}$. Then we have

$$I(t,h) = E[\exp(it^T Q)] = E \left[ 1 - F(h, \ldots, h|X) + F(-h, \ldots, -h|X) \exp \left\{ \sum_{r=0}^{m} \prod_{t=1}^{m} q_{v_t}(t)^{r} g_{v_t}(X) \right\} \right]$$

$$+ \int_X \exp \left\{ \sum_{r=0}^{m} \prod_{t=1}^{m} \left[ G(-u_t/h)^r q_{v_t}(t)^r g_{v_t}(X) \right] f(u(x)) \, du \, dP(x) \right\} = I_1(t,h) + I_2(t,h),$$

where $X$ is the support of $\mathcal{X}$, $\nu = \nu_1 + \nu_2 + \cdots + \nu_m, I_h = I_h(1) + I_h(2) + \cdots + I_h(m), I_h = (-\infty, -h) \otimes (-h, h) \otimes \cdots \otimes (-h, h)$. Applying a two-term Taylor expansion, we obtain that for a sufficiently small $h$,

$$|I_1(t,h)| \leq 1 - E[F(0, \ldots, 0|X) + F(0, \ldots, 0|X) + \cdots + F(0, \ldots, 0|X)] h.$$

where $F(0, \ldots, 0) = \partial F(0, \ldots, 0|X)/\partial u_t$. Let

$$I_2(t,h) = \int_X \int_{X} \int_{X} \int_{X} \exp \left\{ \sum_{r=0}^{m} \prod_{t=1}^{m} \left[ G(-u_t/h)^r q_{v_t}(t)^r g_{v_t}(X) \right] f(u(x)) \, du \, dP(x) \right\}$$

By assumption A4(c), and arguments similar as (A.10) and (A.11) of Whang (2006), there exists a constant $C_1 \in (0, 1)$ such that

$$|I_2(t,h)| < C_1 E \left\{ \sum_{i=1}^{m} F(0, \ldots, 0|X) \right\} h.$$

On the other hand, we have

$$|I_2(t,h)| = O(h^2),$$

which combined with (18) and (19) gives (17). □

**Proof of Theorem 3.** Similar to Whang (2006), using Lemma 2 and the Edgeworth expansion for $I_h(\beta_0)$, we have, for any $c > 0$,

$$P(h(\beta_0) \leq c) = P(\chi^2 \leq c) - n^{-1} \text{tr}(A)p^{-1} c g_\nu(c) + O(n^{-2}) + o(nh^2),$$

where $g_\nu(\cdot)$ denotes the density function of $\chi^2(p)$ distribution, and

$$\text{tr}(A) = n^2 x^T x + 1/2 x^{ijkl} - 1/3 x^{klm} x^{klm}.$$

By definition (8) and Lemma 1, we have

$$n^2 x^T x = n^2 (E X)^T V_n^{-1} (E X) = (nh^2)^{l^2} C_n^2 (\xi^T \Sigma \xi) + o(n^2 h^2),$$

where $\xi = E(X^{(r-1)}|0|X)$. By assumption A3,

$$x^{ijkl} < \infty, \quad x^{klm} < \infty.$$

Hence, from (20) and (21) and $\sup_p n h^r < \infty$, we have

$$P(h(\beta_0) \leq c_1) = \alpha - n^{-1} (nh^2)^{l^2} C_n^2 (\xi^T \Sigma \xi) + O(1)) \, p^{-1} c_1 g_p(\alpha) + o(n^{-1} + n h^2) = \alpha + O(n^{-1}).$$

□

**Proof of Corollary 1.** Let $\lambda = \lambda(\xi_0, \beta(2))$, and $Z_{\alpha}^{(2)}(\xi_0, \beta(2)) = \partial Z_{\alpha}(\xi_0, \beta(2))/\partial \beta(2)$, then $\lambda$, $\beta(2)$ satisfy

$$Q_{\alpha}(\xi_0, \beta(2), \lambda) = \sum_{i=1}^{n} \frac{Z_{\alpha}(\xi_0, \beta(2))}{1 + \lambda Z_{\alpha}(\xi_0, \beta(2))} = 0, \quad \frac{Z_{\alpha}(\xi_0, \beta(2))}{1 + \lambda Z_{\alpha}(\xi_0, \beta(2))} = 0.$$

Expanding $Q_{\alpha}(\xi_0, \beta(2), \lambda)$, and $Q_{\alpha}(\xi_0, \beta(2), \lambda)$ at $(\xi_0, \beta_0(2), 0)$, we have

$$\lambda = P_{\lambda} \xi_0 \Sigma_{\alpha}^{-1} \Sigma_{\alpha}^{-1} \xi_0 + o_p(n^{-1/2}) + h^r$$

and

$$\beta(2) - \beta_0(2) = (\tilde{Z}^{(2)} r \Sigma_{\alpha}^{-1} \tilde{Z}^{(2)} r) - \tilde{Z}^{(2)} r \Sigma_{\alpha}^{-1} \tilde{Z} + o_p(n^{-1/2}),$$
where $P_{\sum_{i=1}^{n}Z_{i}^{2}} = \sum_{i=1}^{n}Z_{i}^{2}(Z_{i}^{2}) + \sum_{i=1}^{n}Z_{i}^{2}$, $\bar{Z} = \sum_{i=1}^{n}Z_{i}Z_{i}^{T}$ and $\bar{Z}^{2} = \sum_{i=1}^{n}Z_{i}^{2}(Z_{i}^{2})$. Note that

$$l_{h}(\lambda_{0}) = l_{h}(\lambda_{0}, \beta_{0}) = 2 \sum_{i=1}^{n} \log(1 + \lambda_{i}Z_{i}(\lambda_{0}, \beta_{0})) = 2 \sum_{i=1}^{n} \lambda_{i}Z_{i}(\lambda_{0}, \beta_{0}) - \sum_{i=1}^{n} \lambda_{i}^{2}Z_{i}(\lambda_{0}, \beta_{0})^{2} + o_{p}(1).$$

Similar to Theorem 2, $\sum_{i=1}^{n}Z_{i}^{2} \rightarrow N(0, I)$ and that $\sum_{i=1}^{n}Z_{i}^{2} \rightarrow N(0, I)$ is symmetric and idempotent, with trace equal to $p - q$. Because $\lambda(\bar{\beta}_{SE}) = 0$, thus $l_{h}(\lambda, \beta) = l_{h}(\bar{\beta}_{SE}) = 0$. Hence the empirical likelihood ratio statistic $l_{h}(\lambda_{0})$ converges to $\chi^{2}_{q}$. \(\Box\)

**Proof of Theorem 4.** This proof is similar to Whang's proof of Theorem 4 and thus is skipped. \(\Box\)

**References**


