Smoothed empirical likelihood for quantile regression models with response data missing at random

Shuanghua Luo¹,² · Changlin Mei¹ · Cheng-yi Zhang²

Abstract This paper studies smoothed quantile linear regression models with response data missing at random. Three smoothed quantile empirical likelihood ratios are proposed first and shown to be asymptotically Chi-squared. Then, the confidence intervals for the regression coefficients are constructed without the estimation of the asymptotic covariance. Furthermore, a class of estimators for the regression parameter is presented to derive its asymptotic distribution. Simulation studies are conducted to assess the finite sample performance. Finally, a real-world data set is analyzed to illustrate the effectiveness of the proposed methods.

Keywords Quantile regression · Smoothed empirical likelihood · Missing at random · Confidence interval

Mathematics Subject Classification 62G05 · 62G20 · 60G42

Shuanghua Luo: Supported by the Science Foundation of the Education Department of Shaanxi Province of China (14JK1305).
Changlin Mei: Supported by the National Natural Science Foundations of China (11271296).
Cheng-yi Zhang: Supported by the National Natural Science Foundations of China (11201362 and 11271297) and the Natural Science Foundations of Shaanxi Province of China (2016JM1009).

✉ Shuanghua Luo
iwantflyluo@163.com

¹ School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an 710049, Shaanxi, China
² School of Science, Xi’an Polytechnic University, Xi’an 710048, Shaanxi, China
1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression (QR) has been an indispensable and versatile tool not only for statistical research but also for many wide applications in economics, finance, biology, medicine, and many other disciplines and attracted considerable attention, resulting in numerous papers (e.g., see Cai and Xu 2008; Cai and Xiao 2012; Chen et al. 2015; Koenker 2005; Lv and Li 2013a; Sherwood et al. 2013; Wei et al. 2012; Whang 2006) devoted to various theoretical extensions of this significant topic.

Why does quantile regression attract many researchers’ attention and have such wide application? One significant reason is that compared to the mean model quantile regression is able to directly estimate the effects of the covariates at different quantiles other than the center of the distribution given by traditional least square regression (LSR) methods, such that quantile regression estimator has an important role in characterizing the entire conditional distribution of a dependent variable given regressors and the robustness property to outlier observations (Whang 2006). Therefore, QR is less sensitive due to its $L_1$ norm and more robust to outliers.

Despite significant theoretical advances and a rapidly growing QR literature, only scant attention has been paid to QR analysis when the data samples contain missing values. This is unfortunate, because in many applications, missing data are a commonplace, and substantial bias can result if one neglects this problem. In fact, response variables are usually missing due to various reasons, such as loss of information caused by uncontrollable factors, unwillingness of some sampled units to supply the desired information, failure on the part of investigators to gather correct information, and so forth. Actually, missing of responses is very common in market research surveys, opinion polls, and many scientific experiments.

Date back to the early 1950s, spurred by advances in computer technology that made previously computational intensive numerical calculations a simple matter, a lot of literature on statistical analysis of data with missing values has flourished, see Aerts et al. (2002), Chen (2014), Horvitz and Thompson (1952), Ibrahim (1990), Little and Rubin (2014), Lipsitz et al. (1998), Robins et al. (1994, 1995), Rubin (1987), Schafer and Graham (2002), Verbeke and Molenberghs (2000), Wang and Rao (2002), Wang et al. (2004), and Wang and Sun (2007) and, hence, proposed several different kinds of methods, such as complete-case (CC) analysis method (Little and Rubin 2014), imputation method (Rubin 1987; Lipsitz et al. 1998; Aerts et al. 2002), inverse probability weighted method (IPW) (Horvitz and Thompson 1952; Robins et al. 1994, 1995), and likelihood based method (Schafer and Graham 2002; Verbeke and Molenberghs 2000; Ibrahim 1990) to handle the missing data problem. Among these methods, imputation becomes the most popular and effective method for managing missing data under MAR. The concept of imputation is to replace the missing data by an appropriate value and then to analyze the data using the standard method as if they were complete. There are many imputation methods for the missing response, see, for example, kernel regression imputation (Cheng 1994; Wang and Rao 2002), linear regression imputation (Wang and Rao 2001; Xue 2009a,b), ratio imputation (Rao 1996), and semi-parametric imputation (Wang et al. 2004).
Although missing data analysis has a long history in statistics, most literature on quantile regression does not contain missing data problems. It is the recent years that quantile regression under missing data attracts researchers’ considerable attention. Yoon (2010) proposed an imputation method where imputed values are drawn from the conditional quantile function of the response for which data are incomplete, but his method is valid only under independent and identically distributed (i.i.d.) errors. Wei et al. (2012) developed an iterative imputation procedure for missing covariates in a linear QR model that is valid under non-i.i.d. error terms. Lv and Li (2013a) discussed smoothed empirical likelihood analysis with missing response in partially linear quantile regression. Sherwood et al. (2013) recently considered an inverse probability weighting QR approach for analyzing health care cost data when the covariates are MAR. Sun et al. (2012) considered QR for competing risk data when the failure type was missing. Chen et al. (2015) discussed efficient QR analysis with missing observations. Kim (2013) proposed imputation methods for quantile estimation under missing at random.

On the other hand, empirical likelihood (EL) method, introduced by Owen (1988) and Owen (1990), has many advantages over normal approximation methods for constructing confidence intervals. One is that it produces confidence intervals and regions whose shape and orientation are determined entirely by the data; the second is that empirical likelihood regions are range preserving and transformation respecting. Many authors have used this method for linear, nonparametric, and semiparametric regression models. About the quantile regression model, Chen and Hall (1993) discussed EL confidence intervals for the population quantiles (no covariates); Wang and Zhu (2011) considered EL for quantile regression models with longitudinal data; Tang and Qin (2012) discussed EL approach for estimating equations with missing data; Chen (2014) investigates empirical likelihood estimator based on parametric quantile regression imputation methods with missing data, and several other authors considered EL for censored survival data, see Zhao and Chen (2008). In these studies on EL method, above some very good theoretical results are established for quantile regression models. However, EL estimation equations based on quantile regression models are not differentiable at parameter points, such that the estimation fails to achieve the higher order accuracy. To reach this asymptotic refinements, some researchers in Lv and Li (2013b), Otsu (2008), Wang et al. (2009), and Whang (2006) do not directly apply the empirical likelihood method to QR models, but construct some smooth functions of sample moments of QR models that the empirical likelihood theory requires. More specifically, since the estimating equations for the standard QR estimator are not smooth, Lv and Li (2013b) discussed smoothed EL confidence intervals for quantile regression parameters with auxiliary information; Whang (2006) proposed a smoothed EL to estimate the parameters of quantile regression model and to construct confidence regions of the model parameters; Otsu (2008) focused on the first-order approximation of a smoothed conditional EL approach for QR models.

Based on the approaches of Whang (2006) and Xue (2009b), we studied in this paper smoothed empirical likelihood methods for quantile regression model with response data at random missing under independent distributed errors. The main contribution of this paper can be summarized as follows:
A smoothed quantile empirical likelihood (SQEL) method is proposed for analysis for quantile regression model with response data at random missing. This method has the following four advantages: (1) since it does not need an adjustment and thus avoids estimating an unknown adjustment factor, it is especially attractive when the adjustment factor is difficult to estimate efficiently; (2) this method overcomes the difficulty in selecting the bandwidth such that the existing data-driven algorithms can be used to select an optimal bandwidth; (3) since different quantiles are used in the imputation instead of actually observed responses or means, SQEL method is less sensitive to outliers; (4) the smooth estimating equations that we adopt enables us to establish high-order properties of the estimator. Thus, the SQEL method is more effective than the quantile empirical likelihood (QEL) method.

A class of SQEL ratios, including smoothed complete-case quantile empirical likelihood (SCCQEL) ratio, smoothed weighted quantile empirical likelihood (SWQEL) ratio, and smoothed imputed quantile empirical likelihood (SIQEL) ratio for the regression parameter are proposed and proved that they are asymptotically $\chi^2$ distributed.

A class of SQEL estimators for the regression parameter are constructed, such that this class of estimators is asymptotically normal. These results obtained in this paper can be used directly to construct the confidence intervals of the regression parameters.

The rest of this paper is organized as follows. A class of SQEL ratios and estimators for the regression parameters $\beta$ are constructed for the quantile linear regression model with missing response data in Sect. 2. Section 3 studies the asymptotic property of the proposed estimators. A simulation study in Sect. 4 demonstrates the finite-sample performance of the proposed method. An application to a real data set illustrates effectiveness of our approach in Sect. 5. The proofs of the main results are given in Sect. 6.

2 Smooted quantile empirical likelihood (SQEL) method with missing response

In this section, we consider the following quantile linear regression model that introduced by Koenker and Bassett (1978):

$$Y_i = X_i^T \beta_\tau + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

(2.1)

where $Y_i$ is the response, $X_i$ is a $d \times 1$ vector of covariates, and $\beta_\tau$ is a $d \times 1$ vector of unknown regression coefficients. Hereafter, whenever there is no confusion, we will simply write $\beta_\tau$ as $\beta$, $\varepsilon_i$ are random error satisfying $P(\varepsilon_i < 0|X_i) = \tau$ for any $i$, where $\tau \in (0, 1)$ is the quantile level of interest. However, different from the model in Koenker and Bassett (1978), we focus on a special case of the model (2.1), where some $Y$ values in a sample of size $n$ may be missing and $X$ is observed completely. That is, the model (2.1) has an incomplete sample $\{X_i; Y_i; \delta_i\}, 1 \leq i \leq n$, where all the $X_i$s are observed completely, $\delta_i = 0$ if $Y_i$ is missing and $\delta_i = 1$, otherwise. Throughout this paper, we assume that $Y$ is missing at random (MAR). The MAR assumption
Smoothed empirical likelihood for quantile regression models

(Cheng 1994) implies that $\delta$ and $Y$ are conditionally independent given $X$, that is, $P(\delta = 1|X; Y) = P(\delta = 1|X)$ denoted by $p(X)$. MAR is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, for details, see Little and Rubin (2014) and Cheng (1994).

2.1 Smoothed complete-case quantile empirical likelihood

For the model (2.1), the complete-case quantile regression (CCQR) estimator $\hat{\beta}_Q$ of $\beta$ is defined by:

$$\hat{\beta}_Q = \arg \min_{\beta \in \mathbb{R}^d} n \sum_{i=1}^{n} \rho_{\tau}(Y_i - X_i^T \beta)\delta_i,$$

(2.2)

where $\rho_{\tau}(u) = u(\tau - I(u < 0))$ is called quantile loss function. To obtain the estimator $\hat{\beta}_Q$, we have with the first-order optimal condition that

$$\sum_{i=1}^{n} \delta_i X_i \psi(Y_i, X_i, \beta) = 0,$$

(2.3)

where $\psi(Y_i, X_i, \beta) = I_{\{X_i^T \beta - Y_i > 0\}} - \tau$ is called quantile score function (see Wang and Zhu 2011) and $I(\cdot)$ is the indicator function, and thus,

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i X_i \psi(Y_i, X_i, \beta) = 0.$$

(2.3)

The left side of the equality (2.3) can be approximately thought as the expectation value of the sample: $\delta_i X_i \psi(Y_i, X_i, \beta), \ i = 1, 2, \ldots, n$. As a result, we obtain the estimating equation:

$$E\{\delta_i X_i \psi(Y_i, X_i, \beta)\} = 0.$$

(2.4)

It should be noted that when $Y_i = X_i^T \beta$ for some $i$, the function $\psi(Y_i, X_i, \beta)$ in (2.4) is not differentiable at points $\beta$. This leads to some difficulties in the subsequent higher order asymptotic analysis, since most of theoretical research on the empirical likelihood has based on a smooth function of sample moments.

To achieve the higher order accuracy, based on the method of Whang (2006), we propose a smooth quantile empirical likelihood (SQEL) approach by approximating $\psi(\cdot)$ in (2.4) by a smooth function $\psi_h(\cdot)$. Let $K(\cdot)$ denote a kernel function that is bounded, compactly supported on $[-1, 1]$ and integrated to one. Define $G(x) = \int_{u < x} K(u)du$ and $G_h(x) = G(x/h)$, where $h$ is a positive bandwidth parameter. Then, a smooth function $\psi_h(\cdot)$ is defined by:

$$\psi_h(Y_i, X_i, \beta) = G_h(X_i^T \beta - Y_i) - \tau$$

(2.5)

to approximate $\psi(\cdot)$. Define a set
\[ P^n = \left\{ p = (p_1, p_2, \ldots, p_n)^T \in \mathbb{R}^n : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i}^{n} p_i Z_{hic}(\beta) = 0 \right\}, \]

where \( p = (p_1, p_2, \ldots, p_n)^T \) is a vector of nonnegative numbers adding to unity and \( Z_{hic}(\beta) = \delta_i X_i \psi_h(Y_i, X_i, \beta) \) with \( \psi_h(Y_i, X_i, \beta) \) defined in (2.5). Then, a smoothed complete-case quantile empirical log-likelihood (SCCQEL) ratio function for \( \beta \) is defined as:

\[ \hat{R}_{hc}(\beta) = -2 \max_{p \in P^n} \left\{ \sum_{i=1}^{n} \log(np_i) \right\}. \] (2.7)

According the definition above, the smoothed complete-case quantile empirical log-likelihood (SCCQEL) estimator of \( \beta \) is defined as:

\[ \hat{\beta}_{SCCQEL} = \arg \min_{\beta \in \mathbb{R}^n} \{ \hat{R}_{hc}(\beta) \}. \]

### 2.2 Smoothed weighted quantile empirical likelihood

Xue (2009b) studied the weighted empirical likelihood method for linear models with missing responses. In what follows, we further investigate this method and propose the smoothed weighted quantile empirical likelihood method for the quantile linear regression model with missing response data.

According to the definition of SCCQEL ratio function (2.7) in Sect. 2.1, the smoothed weighted quantile empirical log-likelihood (SWQEL) ratio function for \( \beta \) is defined as:

\[ \hat{R}_{hw}^{*}(\beta) = -2 \max_{p \in P^n} \left\{ \sum_{i=1}^{n} \log(np_i) \right\}, \] (2.8)

where

\[ P^n = \left\{ p = (p_1, p_2, \ldots, p_n)^T \in \mathbb{R}^n : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i}^{n} p_i Z_{hiw}^{*}(\beta) = 0 \right\}, \]

\[ Z_{hiw}^{*}(\beta) = \delta_i p(X_i) X_i \psi_h(Y_i, X_i, \beta), \] (2.10)

\( p(x) = P(\delta = 1 | X = x) \) is called a selection probability function and \( \psi_h(Y_i, X_i, \beta) \) is defined in (2.5).

Note that the selection probability in (2.10) is regarded as known. If the selection probability is unknown, it can be estimated by a kernel smoothing method. Wang et al. (1997), Chen et al. (2015), and Xue (2009b) present an estimator of \( p(x) \) as follows:

\[ \hat{p}(x) = \frac{\sum_{i=1}^{n} \delta_i L((X_i - x)/a)}{\max\{1, \sum_{i=1}^{n} L((X_i - x)/a)\}}, \] (2.11)

\( \delta_i \) Springer
where $L(\cdot)$ is a kernel function in $\mathbb{R}^d$, and $a = a_n$ is a sequence of positive numbers tending to zero as $n \to \infty$, which controls the amount of smoothing used in estimations. Replacing $p(X_i)$ in (2.10) by its estimator $\hat{p}(X_i)$ in (2.11), a estimation of SWQEL ratio function for $\beta$ can be obtained in the following:

$$\hat{R}_{hw}(\beta) = -2 \max_{p \in \mathcal{P}^n} \left\{ \sum_{i=1}^{n} \log(np_i) \right\},$$

(2.12)

where

$$\mathcal{P}^n = \left\{ p = (p_1, p_2, \ldots, p_n)^T \in \mathbb{R}^n : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_{hiw}(\beta) = 0 \right\},$$

(2.13)

$$Z_{hiw}(\beta) = \frac{\delta_i}{\hat{p}(X_i)} X_i \psi_h(Y_i, X_i, \beta),$$

(2.14)

$\hat{p}(X_i)$ and $\psi_h(Y_i, X_i, \beta)$ are defined in (2.11) and (2.5), respectively.

It follows from the estimation (2.12) of SWQEL ratio function for $\beta$, and the smoothed weighted quantile empirical likelihood (SWQEL) estimator $\hat{\beta}_{SWQEL}$ of $\beta$ is defined as:

$$\hat{\beta}_{SWQEL} = \arg \min_{\beta \in \mathbb{R}_n} \{ \hat{R}_{hw}(\beta) \}.$$  

2.3 Smoothed imputed quantile empirical likelihood

For the SCCQEL and the SWQEL, the information contained in the sample data is not explored fully. Since incomplete-case data are discarded in constructing the empirical likelihood ratio, the coverage accuracies of confidence regions are reduced when there are plenty of missing values. To resolve the issue, Xue (2009b) studied imputation method for empirical likelihood of linear model with missing responses and obtained some good results.

In this section, we continue to analyze this method for quantile regression model with missing response data by imputing $Y_i$ by $X_i^T \hat{\beta}_Q$ if $Y_i$ is missing. Now, we introduce the following auxiliary random:

$$Z_{hiI}(\beta) = X_i \psi_h(\hat{Y}_i, X_i, \beta),$$

(2.15)

where $\hat{Y}_i = \frac{\delta_i Y_i}{\hat{p}(X_i)} + (1 - \frac{\delta_i}{\hat{p}(X_i)}) X_i^T \hat{\beta}_Q$, and $\hat{p}(X_i)$ and $\hat{\beta}_Q$ are defined in (2.11) and (2.2), respectively. Thus, a smoothed imputed quantile empirical log-likelihood (SIQEL) ratio function for $\beta$ is defined as:

$$\hat{R}_{hiI}(\beta) = -2 \max_{p \in \mathcal{P}^n} \left\{ \sum_{i=1}^{n} \log(np_i) \right\},$$

(2.16)
where

\[ P^n = \left\{ p = (p_1, p_2, \ldots, p_n)^T \in \mathbb{R}^n : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i} p_i Z_{hi1}(\beta) = 0 \right\} \]

and \( Z_{hi1}(\beta) \) is defined in (2.15).

The ratio is more appropriate than the SWQEL ratio, because it sufficiently uses the information contained in the data. In addition, \( \hat{\beta}_Q \) of \( Z_{hi}(\beta) \) can be substituted by \( \hat{\beta}_{SCQEL} \). Then, the smoothed imputed quantile empirical likelihood (SIQEL) estimator \( \hat{\beta}_{SIQEL} \) of \( \beta \) is defined as:

\[ \hat{\beta}_{SIQEL} = \arg \min_{\beta \in \mathbb{R}^n} \{ \hat{R}_{hi}(\beta) \}. \]

### 3 Asymptotic properties

Some asymptotic properties are proposed and proved in this section including the asymptotic distributions of the SQEL ratios and the asymptotic normality of the SQEL estimators. These asymptotic results are an extension for the QR models with missing response data of Whang’s asymptotic properties (Whang 2006) for non-missing data.

First, the asymptotic distribution of the SQEL ratios \( \hat{R}_{hc}(\beta), \hat{R}_{hw}(\beta), \) and \( \hat{R}_{hi}(\beta) \) is given as follows.

**Theorem 3.1** Suppose that Conditions C1–C8 in the Appendix hold. If \( \beta \) is the true parameter, then \( \hat{R}_{h}(\beta) \overset{L}{\longrightarrow} \chi_d^2 \), where \( \hat{R}_{h}(\beta) \) is taken to be \( \hat{R}_{hc}(\beta), \hat{R}_{hw}(\beta) \) or \( \hat{R}_{hi}(\beta) \), \( \chi_d^2 \) means the Chi-square distribution with \( d \) degrees of freedom, and \( \overset{L}{\longrightarrow} \) represents the convergence in distribution.

**Remark 3.2** Let \( \chi_d^2 (1 - \alpha) \) be the \( 1 - \alpha \) quantile of the \( \chi_d^2 \) for \( 0 < \alpha < 1 \). According to Theorem 3.1, an approximate \( 1 - \alpha \) confidence region for \( \beta \) is defined by:

\[ R_\alpha(\tilde{\beta}) = \{ \tilde{\beta} \mid \hat{R}_{h}(\tilde{\beta}) \leq \chi_d^2 (1 - \alpha) \} \]

Theorem 3.1 can also be used to test the hypothesis \( H_0 : \beta = \beta_0 \). One could reject \( H_0 \) at level \( \alpha \) if \( \hat{R}_{h}(\beta_0) > \chi_d^2 (1 - \alpha) \).

To compare the empirical likelihood method with a normal approximation method, the following theorem gives the asymptotic normality of \( \hat{\beta}_Q \) and \( \hat{\beta}_{SQEL} \), where \( \hat{\beta}_{SQEL} \) is taken to be \( \hat{\beta}_{SCQEL}, \hat{\beta}_{SWQEL} \), and \( \hat{\beta}_{SIQEL} \).

**Theorem 3.3** Suppose that Conditions C1–C8 in the Appendix hold. Then, the following conclusions hold:

1. \( \sqrt{n}(\hat{\beta}_{SQEL} - \hat{\beta}_Q) = \sigma_p(1) \) and \( \sqrt{n}(\hat{\beta} - \beta) \overset{L}{\longrightarrow} N(0, D_1) \), where \( D_1 = A_1^{-1} B_1 A_1^{-1}, A_1 = E\{p(X)f(0|X)XX^T\}, B_1 = \tau(1 - \tau)E\{p(X)XX^T\} \) and \( \beta \) is taken to be \( \hat{\beta}_{SCQEL} \) or \( \hat{\beta}_Q \).
To construct the confidence region of $\beta$, the asymptotic covariance matrix $D_1$ can be estimated by $\hat{D}_1 = \hat{A}_1^{-1} \hat{B}_1 \hat{A}_1^{-1}$, where $\hat{A}_1 = \frac{1}{n} \sum_{i=1}^{n} \delta_i K_h(Y_i - X_i^T \hat{\beta}) X_i X_i^T$ and $\hat{B}_1 = \frac{\tau(1-\tau)}{n} \sum_{i=1}^{n} \delta_i X_i X_i^T$. It can be proved that $\hat{D}_1$ is a consistent estimator of $D_1$. Thus, it follows from Theorem 3.3 that

$$
\hat{D}_1^{-1/2} \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, I_d),
$$

where $I_d$ is the identity matrix of order $d$. Furthermore,

$$
(\hat{\beta} - \beta)^T n \hat{D}_1^{-1} (\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \chi^2_d.
$$

(3.1)

By the same method, we can obtain:

$$
(\hat{\beta}_{SWQEL} - \beta)^T n \hat{D}_2^{-1} (\hat{\beta}_{SWQEL} - \beta) \xrightarrow{\mathcal{L}} \chi^2_d,
$$

(3.2)

and

$$
(\hat{\beta}_{SIQEL} - \beta)^T n \hat{D}_3^{-1} (\hat{\beta}_{SIQEL} - \beta) \xrightarrow{\mathcal{L}} \chi^2_d,
$$

(3.3)

where $\hat{D}_2 = \hat{A}_2^{-1} \hat{B}_2 \hat{A}_2^{-1}$ and $\hat{D}_3 = \hat{A}_3^{-1} \hat{B}_3 \hat{A}_3^{-1}$ are the estimators of the asymptotic covariance matrices $D_2$ and $D_3$, respectively. Therefore, the confidence regions of $\beta$ can be constructed by using (3.1), (3.2), and (3.3).

4 Simulation study

A simulation study is carried out to investigate the finite-sample properties of the approach based on SQEL. We consider the following two different models:

Model 1 (homoscedastic) : $Y_i = X_i \beta + \epsilon_i$, $i = 1, 2, \ldots, n$;

Model 2 (heteroscedastic) : $Y_i = X_i \beta + \xi X_i \epsilon_i$, $i = 1, 2, \ldots, n$,

where the variable $X$ is simulated from the $N(0, 1)$, $\epsilon_i \sim N(0, 1)$, $\xi = 0.5$, and $\beta = 1$. In this simulation study, we focus on $\tau = 0.4$, 0.7 and adopt the following three selection probability functions proposed by Wang and Rao (see Wang and Rao 2002).

Case 1 : $p_1(x) = 0.8 + 0.2|x - 1| i f \ |x - 1| \leq 1$, and 0.95 otherwise.

Case 2 : $p_2(x) = 0.9 - 0.2|x - 1| i f \ |x - 1| \leq 4.5$, and 0.1 otherwise.

Case 3 : $p_3(x) = 0.6$ for all $x$.

The average missing rates corresponding to the preceding three cases are approximately 0.09, 0.26, and 0.40, respectively.
In what follows, we would give the details for how to generate the missing data for a given selection probability function \( p(X) \). First, one set of data \((X_i, Y_i)\) from the given regression model was generated, and the value of \( p(X_i) \) was calculated. Then, a Bernoulli trial with the probability \( p(X_i) \) for the outcome \( \delta_i = 1 \) was performed. If the trial outcome is \( \delta_i = 1 \), the corresponding \( Y_i \) is observed and reserved; otherwise, \( Y_i \) is missing for each of the three cases. By repeating the above procedure for \( n \) times, the missing data \( \{(X_i, Y_i, \delta_i = 0), i = 1, 2, \ldots, n\} \) were obtained.

For each of the three cases, we generated 2000 Monte Carlo random samples of size \( n = 50, 100, \) and \( 150 \). The kernel functions selected were the same as the ones in Xue (2009a), i.e., \( K(x) = 0.75(1 - x^2)I_{|x| \leq 1} \) and \( L(x) = 0.5I_{|x| \leq 1} \), where \( I_{\cdot} \) is the indicator function. It follows that we used the cross-validation method to select the optimal bandwidths \( a_{\text{opt}} \) and \( h_{\text{opt}} \). The cross-validation criterion are given by

\[
CV(a) = \frac{1}{n} \sum_{i=1}^{n} (\delta_i - \hat{p}_{-i}(X_i))^2 \\
CV(h) = \frac{1}{n} \sum_{i=1}^{n} \delta_i (Y_i - X_i^T \hat{\beta}_{-i})^2,
\]

respectively, where \( \hat{p}_{-i}(X_i) \) is the Nadaraya–Watson estimator of \( p(X_i) \) that is computed when the \( i \)th observation \( X_i \) is deleted, and \( \hat{\beta}_{-i} \) is the SCCQEL estimator of \( \beta \) that is computed when the \( i \)th observation \( X_i \) is deleted.

Now, we discuss the confidence intervals of \( \beta \). We used five methods, including SCCQEL, SWQEL, SIQEL, and SQELW (SQEL without missing data) which came

| \( p(x) \) | \( \tau \) | \( n \) | SQEL | NA | SQELW |
|---|---|---|---|---|---|
| \( p_1(x) \) | 0.4 | 50 | 0.8874 | 0.9012 | 0.9035 | 0.8821 | 0.8794 | 0.8614 |
| & | 100 | 0.6571 | 0.7068 | 0.7078 | 0.7041 | 0.6541 | 0.6445 |
| & | 150 | 0.5089 | 0.5816 | 0.5819 | 0.5110 | 0.5081 | 0.5062 |
| & | 0.7 | 50 | 0.8915 | 0.9145 | 0.9198 | 0.8919 | 0.8906 | 0.8902 |
| & | 100 | 0.7024 | 0.7423 | 0.7389 | 0.7086 | 0.7018 | 0.7016 |
| & | 150 | 0.5691 | 0.5716 | 0.5746 | 0.5635 | 0.5689 | 0.5681 |
| \( p_2(x) \) | 0.4 | 50 | 0.9248 | 0.9048 | 0.8915 | 0.8918 | 0.9021 | 0.8916 |
| & | 100 | 0.7245 | 0.7021 | 0.6946 | 0.6949 | 0.7135 | 0.6947 |
| & | 150 | 0.5843 | 0.5576 | 0.5298 | 0.5301 | 0.5839 | 0.5298 |
| & | 0.7 | 50 | 0.9347 | 0.9175 | 0.9098 | 0.9099 | 0.9321 | 0.9097 |
| & | 100 | 0.7672 | 0.7148 | 0.6869 | 0.6870 | 0.7451 | 0.6867 |
| & | 150 | 0.6514 | 0.5663 | 0.5132 | 0.5133 | 0.6423 | 0.5132 |
| \( p_3(x) \) | 0.4 | 50 | 0.8873 | 0.8554 | 0.8123 | 0.8122 | 0.8742 | 0.8124 |
| & | 100 | 0.6517 | 0.6423 | 0.6147 | 0.6148 | 0.6418 | 0.6146 |
| & | 150 | 0.5248 | 0.5104 | 0.5012 | 0.5015 | 0.5189 | 0.5013 |
| & | 0.7 | 50 | 0.8997 | 0.8874 | 0.8698 | 0.8701 | 0.8869 | 0.8697 |
| & | 100 | 0.7048 | 0.6798 | 0.6589 | 0.6592 | 0.6724 | 0.6588 |
| & | 150 | 0.5683 | 0.5589 | 0.5109 | 0.5114 | 0.5582 | 0.5110 |
from earlier work of Whang (2006) and normal approximation (NA) based on Theorem 3.3. For convenience, in the rest of this paper, NA(\hat{\beta}_{\text{SIQEL}}) and NA(\hat{\beta}_Q) denote the corresponding normal approximation confidence intervals for \hat{\beta}_{\text{SIQEL}} and \hat{\beta}_Q. The average lengths of confidence intervals and their corresponding empirical coverage probabilities, with a nominal level 1 − \alpha = 0.95 and \tau = 0.4, 0.7, were computed with 2000 simulation runs. The results are reported in Tables 1, 2, 3, and 4.

From Tables 1, 2, 3, and 4, we can obtain the following results. First, under Case 1, SIQEL has slightly longer interval lengths, but higher coverage probabilities, than the other four methods. For Cases 2 and 3, SIQEL and SQELW have nearly equal interval lengths and coverage accuracies, while SIQEL performs better than the other three methods, because its confidence intervals have uniformly shorter average lengths and higher coverage probabilities. This shows that quantile regression imputation is necessary when the missing rate is large. Secondly, both SCCQEL and SWQEL have slightly longer interval lengths but higher coverage probabilities than NA(\hat{\beta}_Q). In addition, the confidence intervals obtained by NA(\hat{\beta}_{\text{SIQEL}}) have uniformly shorter average lengths and higher coverage probabilities than NA(\hat{\beta}_Q). Third, all the interval lengths decrease, while the empirical coverage probabilities increase for every fixed missing rate as n increases. Observably, the missing rate also affects the interval

| p(x) | \tau | n | SQEL | SCCQEL | SWQEL | SIQEL | NA(\hat{\beta}_{\text{SIQEL}}) | NA(\hat{\beta}_Q) | SQELW |
|------|------|---|------|--------|-------|-------|-----------------------------|----------------|-------|
| p_1(x) | 0.4  | 50 | 0.8898 | 0.8924 | 0.9008 | 0.9004 | 0.8875 | 0.9007 |
|       |      | 100| 0.9076 | 0.9146 | 0.9204 | 0.9202 | 0.9053 | 0.9203 |
|       |      | 150| 0.9186 | 0.9208 | 0.9312 | 0.9304 | 0.9176 | 0.9311 |
| 0.7   | 50   | 0.8948 | 0.9054 | 0.9146 | 0.9114 | 0.8922 | 0.9145 |
|       |      | 100| 0.9016 | 0.9198 | 0.9298 | 0.9283 | 0.9011 | 0.9298 |
|       |      | 150| 0.9154 | 0.9241 | 0.9311 | 0.9309 | 0.9123 | 0.9310 |
| p_2(x) | 0.4  | 50 | 0.9006 | 0.9049 | 0.9147 | 0.9131 | 0.9001 | 0.9146 |
|       |      | 100| 0.9106 | 0.9119 | 0.9206 | 0.9192 | 0.9102 | 0.9206 |
|       |      | 150| 0.9189 | 0.9248 | 0.9381 | 0.9342 | 0.9178 | 0.9380 |
| 0.7   | 50   | 0.9042 | 0.9153 | 0.9172 | 0.9161 | 0.9038 | 0.9171 |
|       |      | 100| 0.9114 | 0.9204 | 0.9246 | 0.9234 | 0.9102 | 0.9246 |
|       |      | 150| 0.9224 | 0.9302 | 0.9382 | 0.9362 | 0.9212 | 0.9383 |
| p_3(x) | 0.4  | 50 | 0.9013 | 0.9115 | 0.9191 | 0.9187 | 0.9002 | 0.9190 |
|       |      | 100| 0.9103 | 0.9208 | 0.9276 | 0.9258 | 0.9098 | 0.9276 |
|       |      | 150| 0.9205 | 0.9315 | 0.9386 | 0.9357 | 0.9197 | 0.9386 |
| 0.7   | 50   | 0.9025 | 0.9132 | 0.9204 | 0.9199 | 0.9014 | 0.9204 |
|       |      | 100| 0.9112 | 0.9219 | 0.9286 | 0.9277 | 0.9104 | 0.9285 |
|       |      | 150| 0.9241 | 0.9348 | 0.9409 | 0.9398 | 0.9214 | 0.9408 |
Table 3  Average lengths of the confidence intervals for $\beta$ in Model 2 for different forms of the selection probability function $p(x)$ and different values of sample size $n$ and the quantile level $\tau$ under the nominal level 0.95.

| $p(x)$ | $\tau$ | $n$ | SQEL | SCCQEL | SWQEL | SIQEL | NA(\hat{\beta}_{SIQEL}) | NA(\hat{\beta}_Q) | SQELW |
|--------|--------|-----|------|--------|--------|-------|---------------------|-----------------|--------|
| $p_1(x)$ | 0.4 | 50 | 0.8961 | 0.8969 | 0.9112 | 0.8998 | 0.8957 | 0.8929 |
| | | 100 | 0.6562 | 0.6645 | 0.7158 | 0.6914 | 0.6513 | 0.6548 |
| | | 150 | 0.5515 | 0.5873 | 0.5914 | 0.5719 | 0.5478 | 0.5318 |
| | 0.7 | 50 | 0.9058 | 0.9145 | 0.9368 | 0.9256 | 0.9014 | 0.9102 |
| | | 100 | 0.7094 | 0.7252 | 0.7464 | 0.7146 | 0.7078 | 0.7149 |
| | | 150 | 0.5784 | 0.5795 | 0.5801 | 0.5767 | 0.5756 | 0.5801 |
| $p_2(x)$ | 0.4 | 50 | 0.9136 | 0.9089 | 0.8865 | 0.8875 | 0.9067 | 0.8865 |
| | | 100 | 0.7187 | 0.7046 | 0.6974 | 0.6984 | 0.7024 | 0.6973 |
| | | 150 | 0.5874 | 0.5578 | 0.5314 | 0.5331 | 0.5568 | 0.5314 |
| | 0.7 | 50 | 0.9342 | 0.9145 | 0.9426 | 0.9413 | 0.9412 | 0.9126 |
| | | 100 | 0.7456 | 0.7165 | 0.6874 | 0.6892 | 0.7154 | 0.6875 |
| | | 150 | 0.6528 | 0.5672 | 0.5164 | 0.5161 | 0.5649 | 0.5165 |
| $p_3(x)$ | 0.4 | 50 | 0.8786 | 0.8478 | 0.8149 | 0.8159 | 0.8456 | 0.8149 |
| | | 100 | 0.6452 | 0.6421 | 0.6164 | 0.6173 | 0.6374 | 0.6164 |
| | | 150 | 0.5257 | 0.5142 | 0.5043 | 0.5066 | 0.5114 | 0.5043 |
| | 0.7 | 50 | 0.9014 | 0.8867 | 0.8785 | 0.8812 | 0.8808 | 0.8786 |
| | | 100 | 0.7014 | 0.6879 | 0.6606 | 0.6636 | 0.6749 | 0.6607 |
| | | 150 | 0.5748 | 0.5578 | 0.5148 | 0.5154 | 0.5542 | 0.5147 |

length and coverage probability. Generally, the interval length increases, while the coverage probability decreases as the missing rate increases for every fixed sample size. However, the two values do not change by a large amount for the SIQEL method, because the regression imputation is used in SIQEL. Furthermore, it is also seen that there is not much difference in relative performance of confidence regions and coverage probabilities under the heteroscedastic model, and SIQEL still performs much better than the SCCQEL and SWQEL methods for the heteroscedastic model. This implies that the proposed method is robust.

5 A real data example

In this section, the data originally presented by Engel (1857) are investigated to support the proposition that food expenditure constitutes a declining share of personal income. These data that have not any missing data and consists of 235 budget surveys of the 19th century European working class households. More details of the discussion on these data can be found in Perthel (1975). To use the data set to illustrate our methods, artificial missing data were created by deleting some of the response values at random. Assume that 25% of the response values in these data are missing at random. We
Table 4  Empirical coverage probabilities of the intervals for $\beta$ in Model 2 for different forms of the selection probability function $p(x)$ and different values of sample sizes $n$ and $\tau$ under nominal level 0.95

| $p(x)$ | $\tau$ | $n$ | SCCQEL | SWQEL | SIQEL | NA($\hat{\beta}_{SIQEL}$) | NA($\hat{\beta}_Q$) | SQELW |
|--------|--------|-----|--------|--------|-------|--------------------------|---------------------|-------|
| $p_1(x)$ | 0.4    | 50  | 0.8967 | 0.8987 | 0.9015 | 0.9009 | 0.8946 | 0.9012 |
|        |        | 100 | 0.9108 | 0.9143 | 0.9203 | 0.9197 | 0.9098 | 0.9201 |
|        |        | 150 | 0.9164 | 0.9187 | 0.9284 | 0.9272 | 0.9148 | 0.9282 |
|        | 0.7    | 50  | 0.9014 | 0.9124 | 0.9184 | 0.9162 | 0.8999 | 0.9181 |
|        |        | 100 | 0.9126 | 0.9205 | 0.9267 | 0.9235 | 0.9097 | 0.9265 |
|        |        | 150 | 0.9241 | 0.9283 | 0.9392 | 0.9373 | 0.9205 | 0.9390 |
| $p_2(x)$ | 0.4    | 50  | 0.9024 | 0.9066 | 0.9192 | 0.9175 | 0.9009 | 0.9191 |
|        |        | 100 | 0.9146 | 0.9159 | 0.9204 | 0.9198 | 0.9107 | 0.9205 |
|        |        | 150 | 0.9269 | 0.9299 | 0.9306 | 0.9299 | 0.9242 | 0.9306 |
|        | 0.7    | 50  | 0.9068 | 0.9122 | 0.9212 | 0.9201 | 0.9021 | 0.9211 |
|        |        | 100 | 0.9149 | 0.9243 | 0.9318 | 0.9304 | 0.9115 | 0.9318 |
|        |        | 150 | 0.9276 | 0.9304 | 0.9396 | 0.9365 | 0.9213 | 0.9397 |
| $p_3(x)$ | 0.4    | 50  | 0.9085 | 0.9172 | 0.9237 | 0.9216 | 0.9041 | 0.9236 |
|        |        | 100 | 0.9164 | 0.9248 | 0.9326 | 0.9305 | 0.9134 | 0.9325 |
|        |        | 150 | 0.9288 | 0.9343 | 0.9414 | 0.9397 | 0.9242 | 0.9414 |
|        | 0.7    | 50  | 0.9097 | 0.9152 | 0.9254 | 0.9212 | 0.9047 | 0.9253 |
|        |        | 100 | 0.9189 | 0.9258 | 0.9335 | 0.9304 | 0.9135 | 0.9334 |
|        |        | 150 | 0.9297 | 0.9376 | 0.9416 | 0.9396 | 0.9241 | 0.9415 |

Consider the following linear QR model: $Y_i = \beta_0(\tau) + \beta_1(\tau)X_i, \ i = 1, 2, \ldots, 235$, where $Y$ is the centered annual household food expenditure and $X$ is the centered annual household income in Belgian francs.

We now present the estimator and the 95% confidence interval of $\beta$ based on the proposed SIQEL method and the normal approximation method (NA($\hat{\beta}_{SIQEL}$)) with $\tau = 0.4$ and 0.7. In addition, for comparison, we also give the results based on the SWQEL and SQELW. The results are shown in Table 5. From Table 5, we can obtain the following results. First, the confidence interval obtained by the SIQEL method has shorter confidence interval than that obtained by the SWQEL method and NA($\hat{\beta}_{SIQEL}$). Second, the SQELW and the SIQEL have nearly equal interval lengths, and NA($\hat{\beta}_{SIQEL}$) is slightly longer than that based on the SIQEL method and on the SQELW method. This also verifies the results obtained in the simulation studies.

6 Summary

In this paper, a smoothed quantile empirical likelihood (SQEL) method is proposed for analysis for quantile regression model with response data at random missing. A class of SQEL ratios, including smoothed complete-case quantile empirical likelihood (SCCQEL) ratio, smoothed weighted quantile empirical likelihood (SWQEL) ratio,
Table 5  The estimators and 95% confidence intervals of $\beta$ based on SWQEL, SIQEL, SQELW, and NA($\hat{\beta}_{SIQEL}$) in Engel data analysis

| Methods     | $\tau$ | Estimators | Confidence intervals |
|-------------|--------|------------|----------------------|
| SWQEL       | 0.4    | $\hat{\beta}_0$ | 101.12 (63.48, 126.32) |
|             |        | $\hat{\beta}_1$ | 0.4998 (0.4679, 0.6046) |
|             | 0.7    | $\hat{\beta}_0$ | 78.85 (51.18, 94.13) |
|             |        | $\hat{\beta}_1$ | 0.5984 (0.5242, 0.6948) |
| SIQEL       | 0.4    | $\hat{\beta}_0$ | 101.02 (65.78, 125.45) |
|             |        | $\hat{\beta}_1$ | 0.4996 (0.4798, 0.5974) |
|             | 0.7    | $\hat{\beta}_0$ | 78.08 (52.47, 92.45) |
|             |        | $\hat{\beta}_1$ | 0.5998 (0.5376, 0.6895) |
| SQELW       | 0.4    | $\hat{\beta}_0$ | 101.06 (65.80, 125.15) |
|             |        | $\hat{\beta}_1$ | 0.5012 (0.4797, 0.5972) |
|             | 0.7    | $\hat{\beta}_0$ | 78.22 (52.48, 92.32) |
|             |        | $\hat{\beta}_1$ | 0.5948 (0.5378, 0.6886) |
| NA($\hat{\beta}_{SIQEL}$) | 0.4    | $\hat{\beta}_0$ | 100.98 (63.12, 130.85) |
|             |        | $\hat{\beta}_1$ | 0.4995 (0.4523, 0.6098) |
|             | 0.7    | $\hat{\beta}_0$ | 77.98 (50.02, 99.76) |
|             |        | $\hat{\beta}_1$ | 0.5996 (0.5247, 0.6978) |

and smoothed imputed quantile empirical likelihood (SIQEL) ratio for the regression parameter, are proposed and proved that they are asymptotically $\chi^2$ distributed. In addition, a class of SQEL estimators for the regression parameter are constructed, such that this class of estimators are asymptotically normal. These results obtained in this paper can be used directly to construct the confidence intervals of the regression parameters. The advantages of the smoothed empirical likelihood method are indicated by a simulation study.

While throughout this paper, we assume that the data are MAR, in practice, the missing mechanism is often of an non-ignorable form, meaning that the quantile score function is dependent on unobserved variables. Unfortunately, this makes the non-ignorable missing data case immensely more complex than MAR. It is clearly an important research topic to develop methods for handling non-ignorable missing data in the QR context. This remains to be done.

Acknowledgements  The authors would like to thank the anonymous referees and the editor for their valuable comments and suggestions, which actually stimulated this work. The authors would also like to thank the Editor-in-Chief, Professor Yarema Okhrin for his sincere help.

7 Appendix

Let $g(x)$ be the density function of $X$, and let $f(\cdot|x)$ and $F(\cdot|x)$ denote the density and distribution functions of $\varepsilon$ conditional on $X_i = x$, respectively. Suppose that $c$ is used to denote a positive constant not dependent on $n$, but it may take a different
value for each appearance. Furthermore, let $r$ be an integer. The following conditions are needed for the convenience of the proof and presentation in Sects. 2 and 3.

(C1) \{$(Y_i, X_i) : i = 1, 2, \ldots, n$\} are independent and identically distributed random vectors.

(C2) Both $p(x)$ and $g(x)$ have bounded partial derivatives up to order $r$ almost surely, and $\inf_x p(x) > 0$.

(C3) This condition includes the following three aspects:
   (i) $K(\cdot)$ is a bounded and compactly supported on $[-1, 1]$.
   (ii) $L(\cdot)$ is a kernel function of order $r$, and there exist positive constants $C_1, C_2,$ and $\rho$, such that $C_1 I_{||u|| \leq \rho} \leq L(u) \leq C_2 I_{||u|| \leq \rho}$.
   (iii) For some constant $C_k \neq 0$, $K(\cdot)$ is an $r$th-order kernel, i.e.,
   \[
   \int u^j K(u) du = \begin{cases} 
   1, & \text{if } j = 1, \\
   0, & \text{if } 1 \leq j \leq r - 1, \\
   C_k, & \text{if } j = r.
   \end{cases} \tag{7.1}
   
   (C4) (see Whang 2006) Let $\tilde{G} = (G(u), \ldots, G^{L+1}(u))^T$ for some $L \geq 1$, where $G(u) = \int_{v < u} K(v) dv$. For any $\theta \in R^{L+1}$ satisfying $||\theta|| = 1$, there is a partition of $[-1, 1]$, $-1 = a_0 < a_1 < \ldots < a_{L+1} = 1$, such that $\theta^T \tilde{G}(u)$ is either strictly positive or strictly negative on $(a_{l-1}, a_l)$ for $l = 1, \ldots, L + 1$.

(C5) The positive bandwidth parameter $h$ satisfies $nh^{2r} \to 0$ and $nh / \log n \to \infty$ as $n \to \infty$.

(C6) The matrices $A$ and $B$ defined in Theorem 3.3 are non-singular.

(C7) $P(||X|| > M_n) = o(n^{-1/2})$, where $0 < M_n \to +\infty$ as $n \to +\infty$.

(C8) The bandwidth $a$ satisfies $na^{2d}M_n^{-2d} \to 0$ and $na^{4r} \to 0$, where $r$ is the order of the Kernel $L(\cdot)$ defined in (ii) of (C3).

In what follows, some lemmas are useful for proving the main theorems given in Sect. 3.

**Lemma 7.1** (see Lemma 2 in Xue 2009a) Suppose that Conditions (C2), (ii) of (C3) and (C7) hold. Then,
   \[
   E\{\hat{p}(X_i) - p(X_i)\}^2 = O((nh^d)^{-1}M_n^d) + O(h^{2r}) + o(n^{-1/2})
   
   holds uniformly over $1 \leq i \leq n$.

**Lemma 7.2** Suppose that Conditions C1–C8 hold. Then, as $n \to \infty$,

(1) $E(\partial Z_{hi}(\beta)/\partial \beta) = A + o(1)$;

(2) $EZ_{hi}(\beta)Z_{hi}^T(\beta) = B + o(1).$ \tag{7.2}

where $A = A_1$ and $B = B_1$ when $Z_{hi}(\beta) = Z_{hic}(\beta)$, $A = A_2$ and $B = B_2$ when $Z_{hi}(\beta) = Z_{hiw}(\beta)$, $A = A_2$ and $B = B_3$ when $Z_{hi}(\beta) = Z_{hil}(\beta)$, and $A_1, A_2, B_1, B_2,$ and $B_3$ are defined in Theorem 3.3.
Proof (a) This lemma will be proved first when $Z_{hi}(\beta) = Z_{hic}(\beta)$. Above all, we only prove Eq. (1). Then, Eq. (2) can also be proved similarly. By a change of variable, we have:

$$Z_{hi}(\beta) = Z_{hic}(\beta) = \delta_i X_i \left[ G \left( \frac{-\varepsilon_i}{h} \right) - F(0|X_i) \right]$$

$$= \delta_i X_i \int [F(-hu|X_i) - F(0|X_i)] K(u) du$$

and then,

$$E \frac{\partial Z_{hi}(\beta)}{\partial \beta^T} = E[p(X_i) f(0|X_i) X_i X_i^T] + E \left[ p(X_i) X_i X_i^T \int [f(-hu|X_i) - f(0|X_i)] K(u) du \right]. \quad (7.3)$$

Then, applying Taylor expansion to the second term in the right-hand side of (7.3) can obtain Eq. (1), where $A_1 = E[p(X_i) X_i X_i^T]$. Similarly, applying Taylor expansion to the first term in the right-hand side of the following equality:

$$E Z_{hi}(\beta) Z_{hi}^T(\beta) = 2E \left[ p(X_i) X_i X_i^T \int [F(-hu|X_i) - F(0|X_i)] [G(u) - \tau] K(u) du \right] + \tau (1 - \tau) E\{p(X_i) X_i X_i^T\}.$$ 

Eq. (2) is obtained immediately, where $B_1 = \tau (1 - \tau) E\{p(X_i) X_i X_i^T\}$.

(b) For the case of $Z_{hi}(\beta) = Z_{hiw}(\beta)$ in Lemma 7.2, direct computations yield:

$$Z_{hiw}(\beta) = \frac{\delta_i}{\hat{p}(X_i)} X_i (G_h(X_i^T \beta - Y_i) - \tau)$$

$$= \frac{\delta_i}{p(X_i)} X_i (G_h(X_i^T \beta - Y_i) - \tau) + \frac{p(X_i) - \hat{p}(X_i)}{\hat{p}(X_i) p(X_i)} \delta_i X_i (G_h(X_i^T \beta - Y_i) - \tau)$$

$$= Z_{hiw}^*(\beta) + \frac{p(X_i) - \hat{p}(X_i)}{\hat{p}(X_i) p(X_i)} \delta_i X_i (G_h(X_i^T \beta - Y_i) - \tau). \quad (7.4)$$

Similar to the proof of Theorem 3 in Wong et al. (2009), it follows from Conditions C2–C5 and C8 that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{p}(X_i) p(X_i)} X_i (G_h(X_i^T \beta - Y_i) - \tau) = O_p(1). \quad (7.5)$$

According to Lemma 7.1, $\sup_x |\hat{p}(x) - p(x)| = o_p(1)$. Consequently, (7.5) indicates:

$$Z_{hiw}(\beta) = Z_{hiw}^*(\beta) + o_p(1). \quad (7.6)$$
Thus, we have

\[
Z_{hi}(\beta) = Z_{hiw}(\beta) = \frac{\delta_i}{p(X_i)} X_i (G_h(X_i^T \beta - Y_i) - \tau) + o_p(1)
\]

\[
= \frac{\delta_i}{p(X_i)} X_i \left[ G \left( \frac{-\epsilon_i}{h} \right) - F(0|X_i) \right] + o_p(1)
\]

\[
= \frac{\delta_i}{p(X_i)} \int \left[ F(\cdot - hu|X_i) - F(0|X_i) \right] K(u) du + o_p(1).
\]

Similar to the proof of the case (a), Eq. (1) is obtained immediately, where

\[
A_2 = E[f(0|X_i)X_i^T].
\]

Similarly, we can also obtain the proof of Eq. (2), where

\[
B_2 = \tau (1 - \tau) E[1/p(X_i)X_i^T].
\]

(c) We now prove Lemma 7.2 when \( Z_{hi}(\beta) = Z_{hiw}(\beta) \). Direct computations yield:

\[
X_i^T \beta - \hat{Y}_i = \left( 1 - \frac{\delta_i}{\hat{p}(X_i)} \right) X_i^T (\beta - \hat{\beta}_Q) - \frac{\delta_i}{\hat{p}(X_i)} (Y_i - X_i^T \beta)
\]

\[
= \left( 1 - \frac{\delta_i}{\hat{p}(X_i)} \right) X_i^T (\beta - \hat{\beta}_Q) - \frac{p(X_i) - \hat{p}(X_i)}{\hat{p}(X_i) p(X_i)} \delta_i \epsilon_i - \frac{\delta_i}{p(X_i)} \epsilon_i.
\]

\[ (7.7) \]

According to Lemma 7.1, \( \sup_x |\hat{p}(x) - p(x)| = o_p(1) \). And it is easily shown that:

\[
\left\| \frac{1}{n} \sum_{i=1}^n \left( 1 - \frac{\delta_i}{\hat{p}(X_i)} \right) X_i \right\| = o_p(1).
\]

Then, \( (7.7) \) indicates:

\[
X_i^T \beta - \hat{Y}_i = -\frac{\delta_i}{p(X_i)} (Y_i - X_i^T \beta) + o_p(1).
\]

\[ (7.8) \]

By \( (7.8) \), we can obtain:

\[
Z_{hiw}(\beta) = (G_h(X_i^T \beta - \hat{Y}_i) - \tau) X_i
\]

\[
= (G_h \left( \frac{\delta_i}{p(X_i)} (X_i^T \beta - Y_i) \right) - \tau) X_i + o_p(1)
\]

\[
= X_i^T \left[ G \left( -\frac{\delta_i}{p(X_i)h} \epsilon_i \right) - F(0|X_i) \right]
\]

\[
= X_i \int \left[ F \left( -\frac{\delta_i}{h} \frac{hp(X_i)}{\delta_i} u|X_i \right) - F(0|X_i) \right] K(u) du + o_p(1),
\]

\[ (7.9) \]
and then,
\[
E \frac{\partial Z_{hi1}(\beta)}{\partial \beta^T} = E[f(0|X_i)X_iX_i^T] + E \left\{ X_iX_i^T \int \left[ f \left( -\frac{hp(X_i)}{\delta_i}u|X_i \right) - f(0|X_i) \right] K(u)du \right\}
\]
\[\text{(7.10)}\]

Therefore, similar to the proof of Case (a), we can finish the proof of Eq. (1) for the case of \( Zhi(\beta) = Zhi I(\beta) \), where \( A^2 = E[f(0|X_i)X_iX_i^T] \). In the same way, we can obtain the proof of Eq. (2), where \( B_3 = \tau(1-\tau)E[X_iX_i^T] \). This completes the proof of Lemma 7.2.

\[\square\]

**Lemma 7.3** Suppose that Conditions C1–C8 hold. Then, as \( n \to \infty \), with probability 1,

\[
\begin{align*}
(1) & \quad \frac{1}{n} \sum_{i=1}^{n} Z_{hi}(\beta) = O(d_n); \\
(2) & \quad \frac{1}{n} \sum_{i=1}^{n} Z_{hi}(\beta)Z_{hi}(\beta)^T = B + O(1),
\end{align*}
\]

uniformly in \( \beta \in T_n \equiv \{ \beta : \| \beta - \beta_0 \| \leq n^{-\eta} \} \), where \( \eta > 0 \), \( B = B_1 \) when \( Z_{hi}(\beta) = Z_{hic}(\beta) \), \( B = B_2 \) when \( Z_{hi}(\beta) = Z_{hiw}(\beta) \), \( B = B_3 \) when \( Z_{hi}(\beta) = Z_{hi1}(\beta) \), and \( B_1, B_2, \) and \( B_3 \) are defined in Theorem 3.3.

**Proof** (a) This lemma will be proved first for the case of \( Z_{hi}(\beta) = Z_{hic}(\beta) \). We only prove Eq. (1). Equation (2) can be proved in a similar way. A Taylor expansion yields:

\[
\frac{1}{n} \sum_{i=1}^{n} Z_{hic}(\beta) = \frac{1}{n} \sum_{i=1}^{n} [Z_{hic}(\beta_0) - EZ_{hic}(\beta_0)] + EZ_{hic}(\beta_0) + R_{nc}(\beta),
\]

where \( R_{nc}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial Z_{hic}(\beta_0^*)}{\partial \beta} (\beta - \beta_0) \) and \( \beta^* \) lies between \( \beta \) and \( \beta_0 \). It then follows from Cauchy–Schwartz inequality, triangle inequality and an argument similar to the proof of Lemma 7.2 that:

\[
\sup_{\beta \in T_n} \| R_{nc}(\beta) \| = O(d_n) \text{ a.s.}
\]  \[\text{(7.12)}\]

Therefore, according to (7.11), (7.12), Eq. (1) in Lemma 7.2 and condition C5, it holds that:

\[
\sup_{\beta \in T_n} \left\| \frac{1}{n} \sum_{i=1}^{n} Z_{hic}(\beta) \right\| = O(n^{-1/2}(\log \log n)^{1/2}) + O(h^r) + O(d_n) = O(d_n) \text{ a.s.}
\]
(b) For the case of $Z_{hi}(\beta) = Z_{hiw}(\beta)$ in Lemma 6.3. Using the Taylor expansion, we obtain:

$$\frac{1}{n} \sum_{i=1}^{n} Z_{hiw}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \{Z_{hiw}(\beta_0) - E Z_{hiw}(\beta_0)\} + E Z_{hiw}(\beta_0) + R_{nw}(\beta),$$  \hspace{1cm} (7.13)

where $R_{nw}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial Z_{hiw}(\beta^*)}{\partial \beta} (\beta - \beta_0)$ and $\beta^*$ lies between $\beta$ and $\beta_0$. Similar to the proof of Case (a), the proof of (1)–(2) is obtained coming from (7.6).

(c) For the case $Z_{hi}(\beta) = Z_{hiI}(\beta)$ in Lemma 6.3. A Taylor expansion yields:

$$\frac{1}{n} \sum_{i=1}^{n} Z_{hiI}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \{Z_{hiI}(\beta_0) - E Z_{hiI}(\beta_0)\} + E Z_{hiI}(\beta_0) + R_{I}(\beta),$$  \hspace{1cm} (7.14)

where $R_{I}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial Z_{hiI}(\beta^*)}{\partial \beta} (\beta - \beta_0)$ and $\beta^*$ lies between $\beta$ and $\beta_0$. Using Lemma 7.2 and (7.9), the proof of (1)–(2) is similar to the proof of the case (a). Therefore, we omit the proof. 

\textbf{Proof of Theorem 3.1}

\textit{Proof} By the Lagrange multiplier method, $\hat{R}_h(\beta)$ can be represented as:

$$\hat{R}_h(\beta) = 2 \sum_{i=1}^{n} \log(1 + \lambda^T(\beta)Z_{hi}(\beta)),$$  \hspace{1cm} (7.15)

where $\lambda(\beta)$ is a $d \times 1$ vector given as the solution to

$$\sum_{i=1}^{n} \frac{Z_{hi}(\beta)}{1 + \lambda^T(\beta)Z_{hi}(\beta)} = 0 \hspace{1cm} (7.16)$$

By Lemma 6.2–6.3, and using the same arguments as are used in the proof of (2.14) in Owen (1990), we can show that:

$$\lambda(\beta) = (Z_{hi}(\beta)Z_{hi}^T(\beta))^{-1} \sum_{i=1}^{n} Z_{hi}(\beta) + o_p(n^{-1/2} + h^r) = O_p(n^{-1/2} + h^r).$$  \hspace{1cm} (7.17)

Applying the Taylor expansion to (7.15), and invoking Lemma 6.1–6.2 and (7.16), we get that:

$$\hat{R}_h(\beta) = 2 \sum_{i=1}^{n} [\lambda^T(\beta)Z_{hi}(\beta) - (\lambda^T(\beta)Z_{hi}(\beta))^2/2] + o_p(1)$$

$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{hi}^T(\beta)\right) \left(\frac{1}{n} \sum_{i=1}^{n} Z_{hi}(\beta)Z_{hi}^T(\beta)\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{hi}(\beta)\right) + o_p(1).$$  \hspace{1cm} (7.18)
This together with Lemma 6.2 and Lemma 6.3 implies that \( \hat{R}_h(\beta) \) has asymptotic Chi-square distribution with \( d \) degrees of freedom. □

**Proof of Theorem 3.3**

**Proof** Noting \( Q_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} Z_{hi}(\beta) \), by Conditions C3 and C4 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 \( \hat{\beta}_{SQEL} \in \text{int}(\mathbb{B}) \) with \( \| \hat{\beta}_{SQEL} - \beta \| \leq cn^{-1/3} + v \), for \( c > 0 \) and \( v > 0 \), such that \( \hat{R}_h(\beta) \) attains its minimum value at \( \hat{\beta}_{SQEL} \). Here, \( \hat{\beta}_{SQEL} \) satisfies \( \lambda(\hat{\beta}_{SQEL}) = 0 \) and \( Q_n(\hat{\beta}_{SQEL}) = 0 \). Applying Taylor expansion to \( Q_n(\hat{\beta}_{SQEL}) \) around \( \beta \), we obtain:

\[
\hat{\beta}_{SQEL} - \beta = -\left( \frac{\partial Q_n(\beta)}{\partial \beta} \right)^{-1} Q_n(\beta) + o_p(n^{-1/2}). \tag{7.19}
\]

Let \( G_{ni} = G(\frac{X_i^T \beta - Y_i}{h}) - I(X_i^T \beta - Y_i \leq 0) \) and rearranging terms, we have:

\[
\sqrt{n} Q_n(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ G \left( \frac{X_i^T \beta - Y_i}{h} \right) - \tau \right] X_i \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [I(X_i^T \beta - Y_i \leq 0) - \tau] X_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [G_{ni} X_i - E G_{ni} X_i] + \sqrt{n} E G_{ni} X_i \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [I(X_i^T \beta - Y_i \leq 0) - \tau] X_i + T_1 + T_2.
\]

By Condition C5, we can obtain \( T_1 = o_p(1) \) and \( T_2 = o_p(1) \). According to Lemma 6.2–6.3, Conditions C4–C8, and (7.19), we complete the proof of Theorem 3.3. □

**References**

Aerts, M., Claeskens, G., Hens, N., Molenberghs, G.: Local multiple imputation. Biometrika 89, 375–388 (2002)

Cai, Z., Xiao, Z.: Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. J. Econom. 167, 413–425 (2012)

Cai, Z., Xu, X.: Nonparametric quantile estimations for dynamic smooth coefficient models. J. Am. Stat. Assoc. 103, 1595–1608 (2008)

Chen, S.: Imputation of missing values using quantile regression. PhD Dissertation, Iowa state University (2014)

Chen, S.X., Hall, P.: Smoothed empirical likelihood confidence intervals for quantiles. Ann. Stat. 22, 1166–1181 (1993)

Chen, X., Wan, A.T.K., Zhou, Y.: Efficient quantile regression analysis with missing observations. J. Am. Stat. Assoc. 110, 723–741 (2015)

Cheng, P.E.: Nonparametric estimation of mean functionals with data missing at random. J. Am. Stat. Assoc. 89, 81–87 (1984)

Engel, E.: Die productions and consumtionsver halttnisse des konigreichs sachsen. Stat. Burdes 8, 1–54 (1857)
Smoothed empirical likelihood for quantile regression models… 115

Horvitz, D.G., Thompson, D.J.: A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47, 663–685 (1952)

Ibrahim, J.: Incomplete data in generalized linear models. J. Am. Stat. Assoc. 85, 765–769 (1990)

Kim, S.Y.: Imputation methods for quantile estimation under missing at random. Stat. Interface 6, 369–377 (2013)

Koenker, R.: Quantiles Regression. Cambridge University Press, Cambridge (2005)

Koenker, R., Bassett Jr., G.: Regression quantiles. Econometrica 46, 33–50 (1978)

Lipsitz, S.R., Zhao, L.P., Molenberghs, G.: A semiparametric method of multiple imputation. J. R. Stat. Soc. Ser. B 60, 127–144 (1998)

Little, R.J.A., Rubin, D.B.: Statistical Analysis with Missing Data. Wiley, New York (2014)

Lv, X., Li, R.: Smoothed empirical likelihood analysis of partially linear quantile regression models with missing response variables. Adv. Stat. Anal. 97, 317–347 (2013a)

Lv, X., Li, R.: Smoothed EL confidence intervals for quantile regression parameters with auxiliary information. Stat. Methodol. 15, 46–54 (2013b)

Otsu, T.: Conditional empirical likelihood estimation and inference for quantile regression models. J. Econom. 142, 508–538 (2008)

Owen, A.B.: Empirical likelihood ratio confidence intervals for a single function. Biometrika 75, 237–249 (1988)

Owen, A.B.: Empirical likelihood ratio confidence regions. Ann. Stat. 18, 90–120 (1990)

Perthel, D.: Engels law revisited. Int. Stat. Rev. 43, 211–218 (1975)

Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. Ann. Stat. 22, 300–325 (1994)

Rao, J.N.K.: On variance estimation with imputed survey data. J. Am. Stat. Assoc. 91, 499–502 (1996)

Robins, J.M., Rotnitzky, A., Zhao, L.P.: Estimation of regression coefficients when some regressors are not always observed. J. Am. Stat. Assoc. 89, 846–866 (1994)

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

Rubin, D.: Multiple Imputations for Nonresponse in Surveys. Wiley, New York (1987)

Schaefer, J., Graham, J.: Missing data: our view of the state of the art. Psychol. Methods 2, 77–147 (2002)

Sherwood, B., Wang, L., Zhou, X.: Weighted quantile regression for analyzing health care cost data with missing covariates. Stat. Med. 32, 4967–4979 (2013)

Sun, Y., Wang, Q., Gilbert, P.: Quantile regression for competing risks data with missing cause of failure. Ann. Stat. 32, 703–728 (2012)

Tang, C.Y., Qin, Y.: An efficient empirical likelihood approach for estimating equations with missing data. Biometrika 99, 1001–1007 (2012)

Verbeke, G., Molenberghs, G.: Linear Mixed Models for Longitudinal Data. Springer, New York (2000)

Wang, C.Y., Wang, S., Zhao, L.-P., Ou, S.-T.: Weighted semiparametric estimation in regression analysis with missing covariate data. J. Am. Stat. Assoc. 92, 512–525 (1997)

Wang, Q., Linton, O., Härdle, W.: Semiparametric regression analysis with missing response at random. J. Am. Stat. Assoc. 99, 334–345 (2004)

Wang, Y.G., Shao, Q.X., Zhu, M.: Quantile regression without the curse of unsmoothness. Comput. Stat. Data Anal. 53, 3696–3705 (2009)

Wang, Q., Rao, J.N.K.: Empirical likelihood for linear regression models under imputation for missing responses. Can. J. Stat. 29, 596–608 (2001)

Wang, Q., Rao, N.K.: Empirical likelihood-based inference under imputation for missing response data. Ann. Stat. 30, 896–924 (2002)

Wang, Q., Sun, Z.: Estimation in partially linear models with missing responses at random. J. Multivar. Anal. 98, 1470–1493 (2007)

Wang, H., Zhu, Z.H.: Empirical likelihood for quantile regression models with longitudinal data. J. Stat. Plan. Inference 141, 1603–1615 (2011)

Wei, Y., Ma, Y., Carroll, R.: Multiple imputation in quantile regression. Biometrika 99, 423–438 (2012)

Whang, Y.J.: Smoothed empirical likelihood methods for quantile regression models. Econom. Theory 22, 173–205 (2006)

Wong, H., Guo, S., Chen, M., Wai-Cheung, I.P.: On locally weighted estimation and hypothesis testing on varying coefficient models. J. Stat. Plan. Inference 139, 2933–2951 (2009)

Xue, L.: Empirical likelihood confidence intervals for response mean with data missing at random. Scand. J. Stat. 36, 671–685 (2009)
Xue, L.: Empirical likelihood for linear models with missing responses. J. Multivar. Anal. 100, 1353–1366 (2009)
Yoon, J.: Quantile regression analysis with missing response with applications to inequality measures and data combination. Working paper (2010)
Zhao, Y., Chen, F.: Empirical likelihood inference for censored median regression model via nonparametric kernel estimation. J. Multivar. Anal. 99, 215–231 (2008)