Smoothed Empirical Likelihood Inference for Nonlinear Quantile Regression Models with Missing Response

Honghua Dong, Xiuli Wang*

School of Mathematics and Statistics, Shandong Normal University, Jinan, China
Email: *wxlmath@163.com

Abstract

In this paper, three smoothed empirical log-likelihood ratio functions for the parameters of nonlinear models with missing response are suggested. Under some regular conditions, the corresponding Wilks phenomena are obtained and the confidence regions for the parameter can be constructed easily.

Keywords

1. Introduction

Quantile regression (QR) proposed by Koenker and Bassett [1] has become a popular alternative to least squares method for providing comprehensive description of the response distribution and robustness against heavy-tailed error distributions. Because of these significant advantages, QR has become an effective method for statistical research. There are many literatures on the estimation of quantile regression models; among them, Koenker [2] is a monograph worth studying. As for some papers, see for example Kim [3], Cai and Xu [4], Wu *et al.* [5], Cai and Xiao [6] and among others.

In recent years, quantile regression with missing data has attracted scholars' considerable attention. There are several methods, such as complete-case (CC) analysis method, inverse probability weighted method (IPW) and imputation method to handle the missing data. For example, Wei *et al.* [7] proposed a multiple imputation estimator for parameter estimation in linear QR with missing covariates. Sherwood *et al.* [8] suggested the inverse probability weighted (IPW) method for linear QR when the covariates are missing at random. Chen *et al.* [9]

also examined the estimation of linear QR model based on nonparametric inverse probability weighted, estimating equations projection, and a combination of both when observations are missing at random. Sherwood [10] investigated the variable selection for the additive partially linear quantile regression with missing covariates. Zhao *et al.* [11] studied several IPW estimators for parameters in QR when covariates or responses are subject to missing not at random.

It is well known that empirical likelihood (EL) method, introduced by Owen [12] [13], has many advantages in constructing confidence intervals. For example, it does not need to construct a pivot quantity, and the confidence regions shape and orientation are determined entirely by the data itself. Some scholars have used this method to QR, and some good theoretical results have been obtained under this framework. See for example, Chen and Hall [14], Wang and Zhu [15], Tang and Leng [16], Zhao *et al.* [17], Zhao and Tang [18], Luo and Pang [19], Zhao and Zhou [20], Zhou *et al.* [21]. However, estimation equations based on quantile regression models are not differentiable at parameter points, such that EL method fails to achieve the higher order accuracy. To achieve the higher-order asymptotic refinements, Whang [22] proposed to smooth the estimating equations for the empirical likelihood under the linear QR models. Later, Lv and Li [23] proposed the smoothed empirical likelihood (SEL) for partially linear quantile regression models with missing response, and the SEL statistics for the parameters and the nonparametric part were defined, and the asymptotic Chi-squared distributions were shown. Recently, for the linear QR models with missing response at random, Luo *et al.* [24] proposed three SEL ratios for the regression parameter, and the asymptotical distributions were shown to be standard χ^2 distribution under some conditions. Linear quantile regression models offer a flexible approach in many applications. It is also of considerable interest to investigate nonlinear quantile models. As far as we know, there is little work done for nonlinear quantile models with missing responses at random. Just like mentioned in Koenker [2] and other literature about nonlinear models, the computation of the entire process in the nonlinear case is considerably more challenging than the linear case where the computation task is quite easy. So it is not directly to extend the work of Whang [22] and Luo *et al.* [24] to nonlinear quantile models because of the complexity of nonlinear models with missing responses at random. Therefore, the main purpose of this paper is to develop the smoothed EL inferences on β with missing responses at random. The rest of this paper is organized as follows. In Section 2, the smoothed empirical likelihood ratios for the parametric vector are proposed, and the asymptotic properties of the proposed empirical log-likelihood ratios are investigated in Section 3. Section 4 is the proofs of the main results. Conclusions are given in Section 5.

2. Methodology

In this paper, we consider the nonlinear quantile regression model

$$Y_i = f\left(X_i, \beta_\tau\right) + \varepsilon_i, i = 1, 2, \cdots, n, \tag{2.1}$$

where X_i is a d-dimensional covariate and Y_i is a response variable, β_τ is a p-dimensional parameter vector, and ε_i is an independent and identically distributed random variable, which satisfies $P(\varepsilon_i \leq 0 \mid X_i) = \tau$, where the quantile level $\tau \in (0,1)$. For simplicity, we write β_τ as β throughout this paper. For model (2.1), we focus on the case where all values of X_i are completely observed, but some values of response Y_i are missing. That is, we have the incomplete observations $(\delta_i, X_i, Y_i), i = 1, 2, \cdots, n$ from model (2.1), where δ_i is an indicator variable, when Y_i can be observed, then $\delta_i = 1$, and when Y_i is missing with $\delta_i = 0$. Throughout this paper, we assume that Y_i is missing at random (MAR), and the MAR assumption satisfies

$$P\{\delta_i = 1 \mid Y_i, X_i\} = P\{\delta_i = 1 \mid X_i\} = \pi(X_i). \tag{2.2}$$

The formula (2.2) implies that δ_i is conditionally independent with Y_i for given X_i, and $\pi(X_i)$ is called the propensity score or selection probability function.

2.1. Smoothed Quantile Empirical Likelihood with Complete-Case Data

In Koenker [2], if $(X_i, Y_i), i = 1, 2, \cdots, n$ can be observed, the quantile estimator $\hat{\beta}$ of the parameter β in model (2.1) is obtained by minimizing the following objective function

$$\sum_{i=1}^{n} \rho_\tau (Y_i - f(X_i, \beta)), \tag{2.3}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function and $I(\cdot)$ is the indicator function. $\hat{\beta}$ can be obtained by solving the following equation, which is the optimal condition corresponding to (2.3), *i.e.*

$$\sum_{i=1}^{n} \nabla f_i \psi(Y_i, X_i, \beta) = 0, \tag{2.4}$$

where $\nabla f_i = \left[\partial f(X_i, \beta) / \partial \beta \right]$, $\psi(Y_i, X_i, \beta) = I(f(X_i, \beta) - Y_i \geq 0) - \tau$ is the quantile score function, and $E(\psi(Y_i, X_i, \beta)) = 0$ if β is the true value.

Since some values of response Y_i from model (2.1) are missing, with MAR assumption we can prove that

$$E(\delta_i \nabla f_i \psi(Y_i, X_i, \beta)) = 0. \tag{2.5}$$

So based on complete-case data, the quantile estimator $\hat{\beta}_Q$ of the parameter β is the solution of

$$\sum_{i=1}^{n} \delta_i \nabla f_i \psi(Y_i, X_i, \beta) = 0. \tag{2.6}$$

As pointed out by Whang [22], the function $\psi(Y_i, X_i, \beta)$ in (2.6) is not differentiable at point β. This will cause some difficulties in higher-order asymptotic analysis, since most of the empirical likelihood-based research is based on a smooth function of sample moments. Then following Whang [22], let $K(\cdot)$ denote a bounded kernel function that is compactly supported on $[-1,1]$ and

integrated to one, define $G(x) = \int_{u<x} K(u)\mathrm{d}u$, $G_h(x) = G(x/h)$, where h is a positive bandwidth, then a smooth function of ψ is defined as

$$\psi_h(Y_i, X_i, \beta) = G_h(f(X_i, \beta) - Y_i) - \tau. \qquad (2.7)$$

It can be proved that $\psi_h(Y_i, X_i, \beta)$ is asymptotically unbiased.

Introducing the auxiliary random vector

$$\eta_{i,C}(\beta) = \delta_i \nabla f_i \psi_h(Y_i, X_i, \beta). \qquad (2.8)$$

According to the above discussion, we know that $E[\eta_{i,C}(\beta)] = 0$, so the smoothed empirical log-likelihood ratio function of parameter β with complete-case data can be defined as

$$R_C(\beta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \eta_{i,C}(\beta) = 0\right\}, \qquad (2.9)$$

where log(.) is the logarithmic fuction based on e. If zero is inside the convex hull of $(\eta_{1,C}(\beta), \cdots, \eta_{n,C}(\beta))$, then a unique value for $R_C(\beta)$ exists. Using the Lagrange multiplier method and some simple calculations, $R_C(\beta)$ can be written as

$$R_C(\beta) = 2\sum_{i=1}^{n}\log\left(1 + \lambda^{\mathrm{T}}\eta_{i,C}(\beta)\right), \qquad (2.10)$$

where λ is a Lagrange multiplier which is determined by

$$\sum_{i=1}^{n}\frac{\eta_{i,C}(\beta)}{1 + \lambda^{\mathrm{T}}\eta_{i,C}(\beta)} = 0. \qquad (2.11)$$

2.2. Smoothed Weighted Quantile Empirical Likelihood

Similar to Section 2.1, we introduce the following auxiliary random vector

$$\eta_{i,W}(\beta) = \frac{\delta_i}{\pi(X_i)}\nabla f_i \psi_h(Y_i, X_i, \beta). \qquad (2.12)$$

Using the MAR assumption, we can prove that $E(\eta_{i,W}(\beta)) = 0$ if β is the true value, thus the smoothed weighted quantile empirical log-likelihood ratio function for β can be defined accordingly.

Since (2.12) contains an unknown function $\pi(X_i)$ which needs to be estimated first. We can use a kernel smoothing method to estimate. Specifically, $\pi(X_i)$ can be defined by

$$\hat{\pi}(X_i) = \frac{\sum_{j=1}^{n} L_a(X_j - X_i)\delta_j}{\sum_{j=1}^{n} L_a(X_j - X_i)}, \qquad (2.13)$$

where $L_a(\cdot) = L(\cdot/a)/a^d$ is the d-dimensional kernel function, and a is the bandwidth.

When the dimension of the covariate X is very high, the nonparametric estimation will encounter with the curse of dimensionality. In this case, a parametric approach might be more feasible for the estimation of $\pi(X_i)$. A commonly used model is the logistic regression given by

$$\pi(X_i, \gamma) = \frac{\exp(\gamma_0 + X_i^{\mathrm{T}} \gamma_1)}{1 + \exp(\gamma_0 + X_i^{\mathrm{T}} \gamma_1)} = \frac{\exp(\Gamma_i^{\mathrm{T}} \gamma)}{1 + \exp(\Gamma_i^{\mathrm{T}} \gamma)}, \tag{2.14}$$

where $\Gamma_i = (1, X_i^{\mathrm{T}})^{\mathrm{T}}$, $\gamma = (\gamma_0, \gamma_1^{\mathrm{T}})^{\mathrm{T}} \in \Theta$ is $(d+1)$-dimensional unknown parameter vector. Here γ can be estimated by maximizing the log-likelihood function

$$L(\gamma) = \sum_{i=1}^{n} \{\delta_i \log \pi(X_i, \gamma) + (1 - \delta_i) \log(1 - \pi(X_i, \gamma))\}. \tag{2.15}$$

Let $\hat{\gamma}$ be the maximum likelihood estimation of γ, then the parameter estimator of $\pi(X_i)$ can be written as $\pi(X_i, \hat{\gamma})$. If the parametric model for $\pi(\cdot)$ is correctly specified, the inverse probability weighted method is consistent and feasible.

For convenience, we use $\hat{\pi}(X_i)$ to represent the estimation of $\pi(X_i)$, it can be the estimator estimated by the parameter method or by nonparametric method. Denote

$$\hat{\eta}_{i,W}(\beta) = \frac{\delta_i}{\hat{\pi}(X_i)} \nabla f_i \psi_h(Y_i, X_i, \beta), \tag{2.16}$$

the smoothed weighted quantile empirical log-likelihood ratio function of parameter β is

$$R_W(\beta) = -2\max\left\{\sum_{i=1}^{n} \log(np_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{\eta}_{i,W}(\beta) = 0\right\}. \tag{2.17}$$

2.3. Smoothed Imputed Quantile Empirical Likelihood

From above discussion, we can see neither approach makes full use of the information contained in the data. As pointed out in Xue [25], discarding the missing data may lead to incorrect conclusion when there are a lot of missing values in the considered data set. To resolve the issue, we first use the nonlinear quantile imputation to impute Y_i by $f(X_i, \hat{\beta}_Q)$, with $\hat{\beta}_Q$ obtained by (2.6), this kind of imputation is also used by Zhao and Tang [18] and Zhou et al. [21]. With the imputed value in hand, and then using the inverse probability weighted technique, we define the final imputed value by

$$\hat{Y}_i = \frac{\delta_i}{\hat{\pi}(X_i)} Y_i + \left(1 - \frac{\delta_i}{\hat{\pi}(X_i)}\right) f(X_i, \hat{\beta}_Q), \tag{2.18}$$

where $\hat{\pi}(X_i)$ is given in Section 2.2. Then the imputation based auxiliary random vector is

$$\hat{\eta}_{i,I}(\beta) = \nabla f_i \psi_h(\hat{Y}_i, X_i, \beta), \tag{2.19}$$

Accordingly, the smoothed imputed quantile empirical log-likelihood ratio function for β is defined as

$$R_I(\beta) = -2\max\left\{\sum_{i=1}^{n} \log(np_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{\eta}_{i,I}(\beta) = 0\right\}. \tag{2.20}$$

The ratio is more appropriate, because it sufficiently uses the information contained in the data.

3. Asymptotic Properties

In this section, we will give the asymptotic distributions for the three smoothed quantile empirical log-likelihood ratios proposed in Section 2.1-2.3. Firstly, we give some symbols and assumptions that needed in proof.

Assuming the probability density function of X is $p(x)$, let $g(\cdot|x)$ and $F(\cdot|x)$ be the density and distribution function of error ε conditional on $X = x$, respectively.

C_1. $\{Y_i, X_i\}, i = 1, \cdots, n$, are independent and identically distributed random vectors.

C_2. Both $\pi(x)$ and $p(x)$ have bounded derivatives up to order r almost surely, and $\inf_x \pi(x) > 0$.

C_3. $L(x)$ is a kernel function of order r, and there is a constant C_1, C_2, ρ, such that $C_1 I(\|x\| \le \rho) \le L(x) \le C_2 I(\|x\| \le \rho)$.

C_4. The kernel function $K(x)$ has bounded and compactly supported on $[-1,1]$, and for a constant $C_K \ne 0$, it satisfies

$$\int u^j K(u)\mathrm{d}u = \begin{cases} 1, & j = 0, \\ 0, & 1 \le j \le r-1, \\ C_K, & j = r. \end{cases}$$

C_5. For $L \ge 1$, let $\tilde{G}(u) = \left(G(u), G^2(u), \cdots, G^{L+1}(u)\right)^{\mathrm{T}}$, where $G(u) = \int_{v<u} K(v)\mathrm{d}v$. For any $\theta \in R^{L+1}$ satisfying $\|\theta\| = 1$, there is a partition on $[-1,1]$: $-1 = b_0 < b_1 < \cdots < b_{L+1} = 1$, such that $\theta^{\mathrm{T}} \tilde{G}(u)$ is either strictly positive or strictly negative on $(b_{l-1}, b_l), l = 1, 2, \cdots, L+1$.

C_6. The bandwidth h satisfies $nh^{2r} \to 0, nh/\log(n) \to \infty$ as $n \to \infty$.

C_7. The matrices $A_i, i = 1, 2$ and $B_i, i = 1, 2, 3$ defined in Lemma 3 of Section 4 are non-singular.

C_8. $P(\|X_n\| > M_n) = o(n^{-1/2})$, where $0 < M_n \to \infty$.

C_9. The bandwidth a satisfies $na^{2d} M_n^{-2d} \to 0, na^{4r} \to 0$.

C_{10}. The maximum likelihood estimation $\hat{\gamma}$ of γ is $\sqrt{n}$-consistent and satisfies the regularity condition of asymptotic normality.

The following Theorem states the asymptotic distribution of $R(\beta)$.

Theorem 1. Suppose that conditions in C_1-C_{10} hold, and β is the true value of the parameter, then

$$R(\beta) \xrightarrow{D} \chi_d^2,$$

where $R(\beta)$ can be $R_C(\beta), R_W(\beta), R_I(\beta)$, χ_d^2 is a chi-square distribution with d degrees of freedom, and $\xrightarrow{D}$ represents convergence in distribution.

According to the above theorem, the confidence region of the parameter β can be constructed. More specifically, for a given α with $0 < \alpha < 1$, let $\chi_d^2(1-\alpha)$ satisfies $P(\chi_d^2 \le \chi_d^2(1-\alpha)) = 1-\alpha$, then the approximate $(1-\alpha)$ confidence region for β can be defined as

$$C_\alpha(\beta) = \left\{ \beta \mid R(\beta) \le \chi_d^2(1-\alpha) \right\}. \tag{3.1}$$

4. Proofs

Before giving the proof of the main theorem, some lemmas are useful for proving the main theorem.

Lemma 1 Suppose conditions C_2, C_3 and C_8 hold, then

$$E\left[\hat{\pi}(X_i) - \pi(X_i)\right]^2 = O\left(\left(na^d\right)^{-1} M_n^d\right) + O\left(a^{2r}\right) + o\left(n^{-1/2}\right).$$

Lemma 1 is the Lemma 2 in Xue [25], so the proof is omitted.

Lemma 2 Suppose that conditions C_2, C_{10} hold and the $\pi(x) = \pi(x, \gamma)$ is correctly specified, then

$$\hat{\gamma} - \gamma = O_p\left(n^{-1/2}\right), \quad \max_{1 \le i \le n}\left|\pi(x, \hat{\gamma}) - \pi(x, \gamma)\right| = o_p(1),$$

where γ is the true value.

Lemma 2 is the lemma A.2 of Tang and Zhao [26], please see the proof details in Tang and Zhao [26].

Lemma 3. Suppose conditions C_1-C_{10} hold, then as $n \to \infty$, we have

1) $E\left[\dfrac{\partial \eta_i(\beta)}{\partial \beta}\right] = A + o(1);$

2) $E\left[\eta_i(\beta)\eta_i^{\mathrm{T}}(\beta)\right] = B + o(1),$
When $\eta_i(\beta) = \eta_{i,C}(\beta)$, $A = A_1 = E\left[\pi(X)g(0 \mid X)\nabla f_1(\nabla f_1)^{\mathrm{T}}\right]$,

$B = B_1 = \tau(1-\tau)E\left[\pi(X)\nabla f_1(\nabla f_1)^{\mathrm{T}}\right];$

When $\eta_i(\beta) = \hat{\eta}_{i,W}(\beta)$, $A = A_2 = E\left[g(0 \mid X)\nabla f_1(\nabla f_1)^{\mathrm{T}}\right]$,

$B = B_2 = \tau(1-\tau)E\left[\dfrac{1}{\pi(X)}\nabla f_1(\nabla f_1)^{\mathrm{T}}\right];$

When $\eta_i(\beta) = \hat{\eta}_{i,I}(\beta)$, $A = A_2 = E\left[g(0 \mid X)\nabla f_1(\nabla f_1)^{\mathrm{T}}\right]$,

$B = B_3 = \tau(1-\tau)E\left[\nabla f_1(\nabla f_1)^{\mathrm{T}}\right].$

Proof a) We first prove the lemma when $\eta_i(\beta) = \eta_{i,C}(\beta)$.
For result 1), By a change of variable, we have

$$\eta_{i,C}(\beta) = \delta_i \nabla f_i \int\left[F(-hu \mid X_i) - F(0 \mid X_i)\right]K(u)\mathrm{d}u,$$

and then

$$E\left[\frac{\partial \eta_{i,C}(\beta)}{\partial \beta^T}\right] = E\left[\pi(X_i)g(0 \mid X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right]$$
$$+ E\left[\pi(X_i)\nabla f_i \nabla f_i^{\mathrm{T}} \int\left[g(-hu \mid X_i) - g(0 \mid X_i)\right]K(u)\mathrm{d}u\right]. \tag{4.1}$$

Obviously, the first term on the right-hand of Equation (4.1) is
$A_1 = E\left[\pi(X_i)g(0 \mid X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right]$, and then applying Taylor expansion to the second term on the right-hand of Equation (4.1) can obtain result 1).

For result 2), noticing that

$$
\begin{aligned}
&E\left[\eta_{i,C}(\beta)\eta_{i,C}^{\mathrm{T}}(\beta)\right] \\
&= \tau(1-\tau)E\left[\pi(X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right] \\
&\quad + 2E\left\{\pi(X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\int\left[F(-hu\mid X_i)-F(0\mid X_i)\right]\left[G(u)-\tau\right]K(u)\mathrm{d}u\right\}.
\end{aligned}
\tag{4.2}
$$

Obviously, the first term on the right-hand of Equation (4.2) is $B_1 = \tau(1-\tau)E\left[\pi(X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right]$, and again applying Taylor expansion to the second term on the right-hand of Equation (4.2) can obtain result 2).

b) When $\eta_i(\beta) = \hat{\eta}_{i,W}(\beta)$, by direct calculation, we can derive

$$
\begin{aligned}
\hat{\eta}_{i,W}(\beta) &= \frac{\delta_i}{\hat{\pi}(X_i)}\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right] \\
&= \frac{\delta_i}{\pi(X_i)}\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right] \\
&\quad + \frac{\pi(X_i)-\hat{\pi}(X_i)}{\hat{\pi}(X_i)\pi(X_i)}\delta_i\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right] \\
&= \eta_{i,W}(\beta) + \frac{\pi(X_i)-\hat{\pi}(X_i)}{\hat{\pi}(X_i)\pi(X_i)}\delta_i\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right].
\end{aligned}
\tag{4.3}
$$

In addition, we can prove

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(X_i)\pi(X_i)}\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right] = O_p(1).
\tag{4.4}
$$

According to Lemma 1 and Lemma 2, we can obtain that $\sup_x\left|\hat{\pi}(x)-\pi(x)\right| = o_p(1)$ for both nonparametric estimator and parameter estimator of $\pi(x)$. Using this result and (4.4), we can derive

$$
\hat{\eta}_{i,W}(\beta) = \frac{\delta_i}{\pi(X_i)}\nabla f_i\left[G_h\left(f(X_i,\beta)-Y_i\right)-\tau\right] + o_p(1).
\tag{4.5}
$$

Further derivation leads to

$$
\begin{aligned}
\hat{\eta}_{i,W}(\beta) &= \frac{\delta_i}{\pi(X_i)}\nabla f_i\left[G(-\varepsilon_i/h)-F(0\mid X_i)\right] + o_p(1) \\
&= \frac{\delta_i}{\pi(X_i)}\nabla f_i\int\left[F(-hu\mid X_i)-F(0\mid X_i)\right]K(u)\mathrm{d}u + o_p(1).
\end{aligned}
\tag{4.6}
$$

Similar to the proof in the case of a), it can be seen that

$$
\begin{aligned}
E\left[\frac{\partial\hat{\eta}_{i,W}(\beta)}{\partial\beta^{\mathrm{T}}}\right] &= E\left[g(0\mid X_i)\nabla f_i\nabla f_i^{\mathrm{T}}\right] \\
&\quad + E\left\{\nabla f_i\nabla f_i^{\mathrm{T}}\int\left[g(-hu\mid X_i)-g(0\mid X_i)\right]K(u)\mathrm{d}u\right\}.
\end{aligned}
\tag{4.7}
$$

Obviously, the first term on the right-hand of the Equation (4.7) is $A_2 = E\left[g(0\mid X_i)\nabla f_i\nabla f_i^{\mathrm{T}}\right]$, and using Taylor expansion to the second term on the right hand of (4.7), and result 1) is proved. Similarly, result 2) is also obtained.

c) When $\eta_i(\beta) = \hat{\eta}_{i,I}(\beta)$, direct calculation yield

$$f(X_i, \beta) - \hat{Y}_i$$

$$= \left(1 - \frac{\delta_i}{\hat{\pi}(X_i)}\right)\left[f(X_i, \beta) - f(X_i, \hat{\beta}_Q)\right] - \frac{\delta_i}{\hat{\pi}(X_i)}\left[Y_i - f(X_i, \beta)\right] \qquad (4.8)$$

$$= \left(1 - \frac{\delta_i}{\hat{\pi}(X_i)}\right)\left[f(X_i, \beta) - f(X_i, \hat{\beta}_Q)\right] - \frac{\pi(X_i) - \hat{\pi}(X_i)}{\hat{\pi}(X_i)\pi(X_i)}\delta_i \varepsilon_i - \frac{\delta_i}{\pi(X_i)}\varepsilon_i.$$

Using Taylor expansion to $f(X_i, \hat{\beta}_Q)$ at β, we have

$$f(X_i, \beta) - \hat{Y}_i = \left(1 - \frac{\delta_i}{\hat{\pi}(X_i)}\right)\nabla f_i^{\mathrm{T}}(\hat{\beta}_Q - \beta)$$

$$- \frac{\pi(X_i) - \hat{\pi}(X_i)}{\hat{\pi}(X_i)\pi(X_i)}\delta_i \varepsilon_i - \frac{\delta_i}{\pi(X_i)}\varepsilon_i + o_p(1).$$

Noticing that $E\left(\frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{\pi(X_i)}\right)\nabla f_i^{\mathrm{T}}\right) = 0$ and combined with

$\hat{\beta}_Q - \beta = O_p(n^{-1/2})$, we can derive

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{\hat{\pi}(X_i)}\right)\nabla f_i^{\mathrm{T}}(\hat{\beta}_Q - \beta) = o_p(1).$$

Using $\sup_x |\hat{\pi}(x) - \pi(x)| = o_p(1)$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(X_i)\pi(X_i)}\varepsilon_i = O_p(1)$, then

we get

$$f(X_i, \beta) - \hat{Y}_i = -\frac{\delta_i}{\pi(X_i)}(Y_i - f(X_i, \beta)) + o_p(1). \qquad (4.9)$$

So we obtain

$$\hat{\eta}_{i,I}(\beta) = \nabla f_i \left[G_h\left(f(X_i, \beta) - \hat{Y}_i\right) - \tau\right]$$

$$= \nabla f_i \left\{G_h\left(\frac{\delta_i}{\pi(X_i)}(f(X_i, \beta) - Y_i)\right) - \tau\right\} + o_p(1)$$

$$= \nabla f_i \left[G_h\left(\frac{\delta_i}{\pi(X_i)h}\varepsilon_i\right) - F(0 \mid X_i)\right] + o_p(1) \qquad (4.10)$$

$$= \nabla f_i \int \left\{F\left(-\frac{h\pi(X_i)}{\delta_i}u \mid X_i\right) - F(0 \mid X_i)\right\}K(u)\,\mathrm{d}u + o_p(1),$$

and then

$$E\left[\frac{\partial \hat{\eta}_{i,I}(\beta)}{\partial \beta^{\mathrm{T}}}\right] = E\left[g(0 \mid X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right]$$

$$+ E\left\{\nabla f_i \nabla f_i^{\mathrm{T}}\int\left[g\left(-\frac{h\pi(X_i)}{\delta_i}u \mid X_i\right) - g(0 \mid X_i)\right]K(u)\,\mathrm{d}u\right\}. \qquad (4.11)$$

Similar to the proof in case a), it is obvious that the first term on the right-hand of Equation (4.11) is $A_2 = E\left[g(0 \mid X_i)\nabla f_i \nabla f_i^{\mathrm{T}}\right]$. Similarly, we can obtain result 2) with $B_3 = \tau(1 - \tau)E\left[\nabla f_i \nabla f_i^{\mathrm{T}}\right]$.

Lemma 4 Suppose conditions C_1-C_{10} hold, then as $n \to \infty$,

1) $\frac{1}{n} \sum_{i=1}^{n} \eta_i(\beta) = O(d_n)$;

2) $\frac{1}{n} \sum_{i=1}^{n} \eta_i(\beta) \eta_i^{\mathrm{T}}(\beta) = B + o(1)$,

uniformly in $\beta \in \Theta_n \equiv \{\beta : \|\beta - \beta_0\| \le d_n\}$, with $d_n = n^{-1/3-\zeta}, 0 < \zeta < 1/6$, where $B = B_1$ when $\eta_i(\beta) = \eta_{i,C}(\beta)$; $B = B_2$ when $\eta_i(\beta) = \hat{\eta}_{i,W}(\beta)$; $B = B_3$ when $\eta_i(\beta) = \hat{\eta}_{i,I}(\beta)$, and B_1, B_2, and B_3 are defined in Lemma 3.

Proof: By Taylor expansion, we derive

$$\frac{1}{n} \sum_{i=1}^{n} \eta_i(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left[\eta_i(\beta_0) - E(\eta_i(\beta_0)) \right] + E(\eta_i(\beta_0)) + R_{nc}(\beta), \qquad (4.12)$$

where $R_{nc}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \eta_i(\beta^*)}{\partial \beta}(\beta - \beta_0)$, and β^* lies between β and β_0.

a) This lemma will be proved first for the case of $\eta_i(\beta) = \eta_{i,C}(\beta)$.

Similar to the proof of Lemma 2 of Whang [22], using Cauchy-Schwartz inequality, triangle inequality and arguments similar to the proof of Lemma 3, we have

$$\sup_{\beta \in \Theta_n} \|R_{nc}(\beta)\| = O(d_n) \quad a.s. \qquad (4.13)$$

with $R_{nc}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \eta_{i,C}(\beta^*)}{\partial \beta}(\beta - \beta_0)$. Therefore, according to (4.12), (4.13), law of iterated logarithm, Lemma 3 and condition C_6, it holds that

$$\sup_{\beta \in \Theta_n} \left\| \frac{1}{n} \sum_{i=1}^{n} \eta_{i,C}(\beta) \right\| = O\left(n^{-1/2} (\log(\log n))^{1/2} \right) + O(h^r) + O(d_n) = O(d_n) \quad a.s.$$

The proof for the first result is completed. The second result can be proved in a similar way, here we omit the details.

b) When $\eta_i(\beta) = \hat{\eta}_{i,W}(\beta), \hat{\eta}_{i,I}(\beta)$, according to the Equations (4.5) and (4.10) respectively, and by the similar arguments for $\eta_i(\beta) = \eta_{i,C}(\beta)$, we can derive the two results, here we omit the details.

Proof of Theorem 1 By the Lagrange multiplier method, $R(\beta)$ can be represented as

$$R(\beta) = 2 \sum_{i=1}^{n} \log \left(1 + \lambda^{\mathrm{T}} \eta_i(\beta) \right), \qquad (4.14)$$

where λ is the solution of the following equation

$$\sum_{i=1}^{n} \frac{\eta_i(\beta)}{1 + \lambda^{\mathrm{T}} \eta_i(\beta)} = 0. \qquad (4.15)$$

Similar to the proof in Owen [13], we can prove that

$$\lambda = \left(\sum_{i=1}^{n} \eta_i(\beta) \eta_i^{\mathrm{T}}(\beta) \right)^{-1} \sum_{i=1}^{n} \eta_i(\beta) + o_p\left(n^{-1/2} + h^r \right) = O_p\left(n^{-1/2} + h^r \right). \qquad (4.16)$$

Using Taylor expansion for (4.14), combined with Lemma 4 and (4.16), we

obtain

$$R(\beta) = 2\sum_{i=1}^{n} \left[\lambda^{\mathrm{T}} \eta_i(\beta) - \left(\lambda^{\mathrm{T}} \eta_i(\beta) \right)^2 \Big/ 2 \right] + o_p(1)$$

$$= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i(\beta) \right)^{\mathrm{T}} \left(\frac{1}{n} \sum_{i=1}^{n} \eta_i(\beta) \eta_i^{\mathrm{T}}(\beta) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i(\beta) \right) + o_p(1). \tag{4.17}$$

When $\eta_i(\beta)$ is $\eta_{i,C}(\beta), \hat{\eta}_{i,W}(\beta)$ and $\hat{\eta}_{i,I}(\beta)$, together with the results of Lemma 3 and Lemma 4 respectively, the asymptotic distribution of the smoothed empirical log-likelihood ratio can be proved to be a chi-square distribution with d degrees of freedom.

5. Conclusion

In this paper, we propose three smoothed empirical log-likelihood ratio functions for nonlinear model parameters with missing responses. We obtain the corresponding Wilks phenomenon under some regular conditions, and can easily construct the confidence interval of the parameters. For the type of data missing, we only consider the case where the covariate data is complete and the response variable data is missing. In addition, there are cases where the covariate data is missing and the response variable data is complete. Therefore, the smooth empirical likelihood inference for nonlinear quantile regression models with missing covariates is also worth studying.

Funding

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Koenker, R. and Bassett, J.G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50. https://doi.org/10.2307/1913643

[2] Koenker, R. (2005) Quantiles Regression. Cambridge University Press, Cambridge.

[3] Kim, M. (2007) Quantile Regression with Varying Coefficients. *The Annals of Statistics*, **35**, 92-108. https://doi.org/10.1214/009053606000000966

[4] Cai, Z. and Xu, X. (2008) Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models. *Journal of the American Statistical Association*, **103**, 1595-1608. https://doi.org/10.1198/016214508000000977

[5] Wu, T., Yu, K. and Yu, Y. (2010) Single-Index Quantile Regression. *Journal of Multivariate Analysis*, **101**, 1607-1621. https://doi.org/10.1016/j.jmva.2010.02.003

[6] Cai, Z. and Xiao, Z. (2012) Semiparametric Quantile Regression Estimation in Dynamic Models with Partially Varying Coefficients. *Journal of Econometrics*, **167**, 413-425. https://doi.org/10.1016/j.jeconom.2011.09.025

[7] Wei, Y., Ma, Y. and Carroll, R. (2012) Multiple Imputation in Quantile Regression. *Biometrika*, **99**, 423-438. https://doi.org/10.1093/biomet/ass007

[8] Sherwood, B., Wang, L. and Zhou, X. (2013) Weighted Quantile Regression for Analyzing Health Care Cost Data with Missing Covariates. *Statistics in Medicine*, **32**, 4967-4979. https://doi.org/10.1002/sim.5883

[9] Chen, X., Wan, A.T.K. and Zhou, Y. (2015) Efficient Quantile Regression Analysis with Missing Observations. *Journal of the American Statistical Association*, **110**, 723-741. https://doi.org/10.1080/01621459.2014.928219

[10] Sherwood, B. (2015) Variable Selection for Additive Partial Linear Quantile Regression with Missing Covariates. *Journal of Multivariate Analysis*, **152**, 206-223. https://doi.org/10.1016/j.jmva.2016.08.009

[11] Zhao, P.Y., Tang, N.S. and Jiang, D.P. (2017) Efficient Inverse Probability Weighting Method for Quantile Regression with Nonignorable Missing Data. *Statistics*, **51**, 363-386. https://doi.org/10.1080/02331888.2016.1268615

[12] Owen, A. (1988) Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**, 237-249. https://doi.org/10.1093/biomet/75.2.237

[13] Owen, A. (1990) Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics*, **18**, 90-120. https://doi.org/10.1214/aos/1176347494

[14] Chen, S.X. and Hall, P. (1993) Smoothed Empirical Likelihood Confidence Intervals for Quantiles. *The Annals of Statistics*, **21**, 1166-1181. https://doi.org/10.1214/aos/1176349256

[15] Wang, H. and Zhu, Z.H. (2011) Empirical Likelihood for Quantile Regression Models with Longitudinal Data. *Journal of Statistical Planning and Inference*, **141**, 1603-1615. https://doi.org/10.1016/j.jspi.2010.11.017

[16] Tang, C. and Leng, C. (2012) An Empirical Likelihood Approach to Quantile Regression with Auxiliary Information. *Statistics and Probability Letters*, **82**, 29-36. https://doi.org/10.1016/j.spl.2011.09.003

[17] Zhao, P., Zhou, X. and Lin, L. (2015) Empirical Likelihood for Composite Quantile Regression Modeling. *Journal of Applied Mathematics and Computing*, **48**, 321-333. https://doi.org/10.1007/s12190-014-0804-3

[18] Zhao, P.X. and Tang, X.R. (2016) Imputation Based Statistical Inference for Partially Linear Quantile Regression Models with Missing Responses. *Metrika*, **79**, 991-1009. https://doi.org/10.1007/s00184-016-0586-8

[19] Luo, S. and Pang, S. (2017) Empirical Likelihood for Quantile Regression Models with Response Data Missing at Random. *AStA Advances in Statistical Analysis*, **15**, 317-330. https://doi.org/10.1515/math-2017-0028

[20] Zhao, P. and Zhou, X. (2018) Robust Empirical Likelihood for Partially Linear Models via Weighted Composite Quantile Regression. *Computational Statistics*, **33**, 659-674. https://doi.org/10.1007/s00180-018-0793-z

[21] Zhou, X.S., Zhao, P.X. and Gai, Y.J. (2022) Imputation Based Empirical Likelihood Inferences for Partially Nonlinear Quantile Regression Models with Missing Responses. *AStA Advances in Statistical Analysis*, **106**, 705-722. https://doi.org/10.1007/s10182-022-00441-z

[22] Whang, Y.J. (2006) Smoothed Empirical Likelihood Methods for Quantile Regression Models. *Economic Theory*, **22**, 173-205. https://doi.org/10.1017/S0266466606060087

[23] Lv, X.F. and Li, R. (2013) Smoothed Empirical Likelihood Analysis of Partially Linear Quantile Regression Models with Missing Response Variables. *AStA Advances*

in Statistical Analysis, **97**, 317-347. https://doi.org/10.1007/s10182-013-0210-4

[24] Luo, S.H., Mei, C.L. and Zhang, C.Y. (2017) Smoothed Empirical Likelihood for Quantile Regression Models with Response Data Missing at Random. *AStA Advances in Statistical Analysis*, **101**, 95-116.
https://doi.org/10.1007/s10182-016-0278-8

[25] Xue, L. (2009) Empirical Likelihood for Linear Models with Missing Responses. *Journal of Multivariate Analysis*, **100**, 1353-1366.
https://doi.org/10.1016/j.jmva.2008.12.009

[26] Tang, N.S. and Zhao, P.Y. (2013) Empirical Likelihood-Based Inference in Nonlinear Regression Models with Missing Responses at Random. *Statistics*, **47**, 1141-1159. https://doi.org/10.1080/02331888.2012.658807