AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(5): 8127–8146.
DOI:10.3934/math.2022452
Received: 16 November 2021
Revised: 22 January 2022
Accepted: 09 February 2022
Published: 24 February 2022

Research article

Robust and efficient estimation for nonlinear model based on composite quantile regression with missing covariates

Qiang Zhao[1], **Chao Zhang**[1], **Jingjing Wu**[2] **and Xiuli Wang**[1,*]

[1] School of Mathematics and Statistics, Shandong Normal University, Jinan 250014, China
[2] Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

* **Correspondence:** E-mail: wxlmath@163.com.

Abstract: In this article, two types of weighted quantile estimators were proposed for nonlinear models with missing covariates. The asymptotic normality of the proposed weighted quantile average estimators was established. We further calculated the optimal weights and derived the asymptotic distributions of the correspondingly resulted optimal weighted quantile estimators. Numerical simulations and a real data analysis were conducted to examine the finite sample performance of the proposed estimators compared with other competitors.

Keywords: nonlinear model; composite quantile estimation; weighted quantile average estimator; missing covariates; inverse probability weighting
Mathematics Subject Classification: 62F12, 62G08

1. Introduction

In recent years, regression analysis is widely used in various fields; for example, logistic regression was used to implement distributed classification of large data sets in Wang, Xu and Wu [1]. Traditional regression analysis is based on the mean, which is easy to calculate and is straightforward to interpret. But mean regression may fail for heavy-tailed error distributions, so a series of new regression methods were proposed. Rank regression and quantile regression are robust estimation methods which are widely used. There are many applications in which several response variables are predicted with a common set of predictors. Zhao, Lian and Ma [2] took the possible correlations among the responses into account, and introduced robust reduced-rank estimator via rank regression. Zhang et al. [3] applied rank regression to the varying-coefficient model and proposed a robust multivariate varying-coefficient model based on rank loss that models the relationships among different responses via reduced-rank regression and penalized variable selection. The above two methods are often used to multivariate regression model. Gong, Xu and Chen [4] proposed a penalized modal regression method for additive models in high dimensional. Quantile regression (QR), as introduced by Koenker and Bassett [5], is also a robust regression and can describe the entire conditional distribution of the response variable given the covariates. Because of these significant advantages, QR has become an effective method for statistical research. It is well known that different quantiles may contain different information of error distributions. Therefore, combining different quantile information could appropriately be a feasible way to improve efficiency. With this idea, Zou and Yuan [6] defined a new loss function which is simply an average of the loss function based on different quantiles, and named the new method as composite quantile regression (CQR). CQR could be considered as a useful extension of the quantile regression. Zhao and Xiao [7] pointed out that simple average (using equal weights) is not an efficient way of using distributional information from different quantile regressions. Koenker [8, 9] proposed a more general approach, which assigns different weights to different quantiles. Jiang et al. [10] extended the research on robust and efficient estimation and model selection in high dimensions to nonlinear models. Unfortunately, when the number of quantiles is large, the calculation is very demanding. Therefore, Bloznelis et al. [11] considered a model-averaged quantile estimator with a computationally cheaper alternative and compared its performance to the composite quantile estimator in both low and high dimensional cases.

Classical regression analysis and related theories are based on completely observed data, while missing data are frequently encountered in almost all research areas, such as psychological sciences and medical studies. In cases of missing data, classical statistical methods such as maximum likelihood estimation (MLE) cannot be applied directly to the corresponding statistical analysis. We know that the complete-case (CC) method, which only uses the fully observed data, can lead to seriously biased parameter estimations when the covariate is not missing completely at random. Yates [12] introduced an imputation method which is widely used to handle missing responses. This method aims to find an appropriate value that to be filled in for each missing data. Then the data with the filled in values can be treated as fully observed data that can be analyzed by classical methods. Xia [13] employ the profile nonlinear least squares estimation based on the weighted imputation method to estimate the unknown parameter and nonparametric function and consider empirical likelihood inferences based on the weighted imputation method for the varying coefficient partially nonlinear model with missing responses. The inverse probability weighted (IPW) method is another frequently used method dated back to Horvitz and Thompson [14] that can be applied to the case of missing covariates. In this method, the inverse of the selection probability is chosen to be the weight assigned to the fully observed data. The missing at random (MAR) assumption, in the sense of Rubin [15], is a common assumption for statistical analysis with missing data. Under the MAR assumption, many approaches for mean regression with missing values were developed to obtain efficient estimators, such as the imputation method proposed by Little and Rubin [16], the IPW method introduced by Robins et al. [17], and likelihood-based methods given by Ibrahim et al. [18]. For a comprehensive review, readers are referred to Qin, Shao and Zhang [19]. It is worth mentioning that IPW method is unbiased under MAR assumption.

However, most of the above methods are built on least squares (LS) estimator which is not robust against outliers. Recently, Sherwood, Wang and Zhou [20]considered a linear QR approach based on IPW with a parametric model for the selection probability when covariates are missing at random, and investigated the variable selection problem with the proposed method. Chen, Wan and Zhou [21] proposed three estimation methods for a linear quantile regression when observations are missing at random, one of which is to use nonparametric IPW. The above three references focused on a given individual quantile. Due to the effectiveness and robustness of the CQR method, Yang and Liu [22] investigated the CQR estimation of linear models with missing covariates by using IPW method. It is worth pointing out that, they used equal weights at different quantiles to construct their CQR estimator for a linear model. Recently, Wang, Song and Zhang [23] proposed an optimal weighted quantile average estimation for parameters in additive partially linear models with missing covariates, and their simulation results verified that the proposed method is an efficient and reliable alternative of both the weighted least squares (WLS) method and the weighted CQR (WCQR) method. So in this paper, applying the idea of Jiang et al. [24] and Wang, Song and Zhang [23], we consider two types of WCQRs for nonlinear models with missing covariates and the proposed methods are demonstrated superior via simulation studies and a real data example.

The rest of this paper is organized as follows. The proposed estimation technique and its theoretical properties are presented in Section 2. Numerical simulation studies are conducted in Section 3 in order to examine the performance of the proposed methods and to justify the derived theoretical results in Section 2. A real data analysis is given in Section 4 to illustrate the implementation of the proposed methods. The regularity conditions and the proofs of those theoretical results are given in Appendix.

2. Methodology

Zhao and Lian [25] studied two weighting schemes to further improve the efficiency of CQR for linear models. And they showed that the two weighting schemes are asymptotically equivalent to each other and always result in more efficient estimators compared with CQR in theory. Now, In order to get a more general approach, we generalize the linear models to the nonlinear models and consider the covariates missing at random. Consider the nonlinear model

$$Y_i = f(X_i,\beta) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where Y_i is an observable response, $X_i = (U_i^T, V_i^T)^T \in R^{q+s}$ is the vector of covariates, β is the p-dimensional vector of unknown parameters, and ε_i is the random error independent of X. Let K be the number of quantiles, for the equally spaced quantiles $\tau_k = \frac{k}{K+1}, k = 1, 2, \ldots, K$. Jiang et al. [24] proposed the weighted composite quantile estimator for β by minimizing

$$l_n(\beta, \mathbf{b}) = \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k})$$

over β and $\mathbf{b} = (b_{\tau_1}, b_{\tau_2}, \ldots, b_{\tau_k})^T$, where $\rho_\tau(t) = t(\tau - I(t < 0))$, and ω_k is the weight which controls the amount of contribution of the τ_k-th quantile regression satisfying $\sum_{k=1}^{K} \omega_k g(b_{\tau_k}) > 0$ with $g(\cdot)$ being the density of ε.

Here we assume some covariates are missing. More specifically, we assume U_i's are all observed while some V_i's are missing. Let $\delta_i = 0$ if V_i is missing, and $\delta_i = 1$ if V_i is observed. Throughout this paper, following Wang, Song and Zhang [23], we assume the following missing mechanism

$$P(\delta_i = 1|Y_i, U_i, V_i) = P(\delta_i = 1|U_i) \triangleq \pi(U_i), \tag{2.2}$$

where $\pi(\cdot)$ is called the selection probability function or the propensity score.

When the selection probability function $\pi(\cdot)$ is known, the IPW estimator of β under missing covariates is defined as

$$(\hat{\mathbf{b}},\hat{\beta}) = \arg\min_{b,\beta} L_n(\pi(U),\beta,\mathbf{b}), \tag{2.3}$$

where $L_n(\pi(U),\beta,\mathbf{b}) = \sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\frac{\delta_i}{\pi(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k})$. However, in reality the selection probability function $\pi(\cdot)$ is usually unknown and needs to be estimated. Next we follow Wang, Song and Zhang [23] and consider estimating $\pi(U_i)$ using both parametric and nonparametric models.

2.1. Estimation of propensity scores

To estimate the propensity scores nonparametrically, we apply nonparametric smoothing techniques. Particularly, we use the Nadaraya-Watson estimator of $\pi(U_i)$ which is defined as

$$\hat{\pi}(U_i) = \frac{\sum_{j=1}^{n}K_h(U_i - U_j)\delta_j}{\sum_{j=1}^{n}K_h(U_i - U_j)}, \tag{2.4}$$

where $K_h(\cdot) = K(\cdot/h)/h^q$ is a q-variate kernel function, h is the bandwidth.

When the dimension of U is high, a fully nonparametric estimation is encountered with the curse of dimensionality. In this case, a parametric approach might be more feasible for the estimation of $\pi(U_i)$ given in (2.2). A commonly used model for (2.2) is the logistic regression given by

$$\pi(U_i,\gamma) = \frac{\exp(\gamma_0 + U_i^T\gamma_1)}{1+\exp(\gamma_0 + U_i^T\gamma_1)} = \frac{\exp(\Gamma_i^T\gamma)}{1+\exp(\Gamma_i^T\gamma)}, \tag{2.5}$$

where $\Gamma_i = (1, U_i^T)^T$ and $\gamma = (\gamma_0,\gamma_1^T)^T \in \Theta$ is an unknown parameter vector with $\Theta \subset R^{q+1}$. Here γ can be estimated by maximizing the log-likelihood function

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

Let $\hat{\gamma}$ be the MLE of γ, then the parametric estimator of $\pi(U_i)$ is denoted by $\pi(U_i,\hat{\gamma})$. If the specified parametric model (2.5) of the selection probability function $\pi(\cdot)$ is valid, then the IPW method is applicable.

2.2. WCQR estimation of regression parameters

In this subsection, we propose two weighting schemes for the WCQR estimation. The first one is based on weighting the quantile loss and the second one is weighting the quantile regression estimator at different levels with details given below. For convenience, we use $\hat{\pi}(U_i)$ for the estimator of $\pi(U_i)$ by either the parametric or nonparametric method.

As in Jiang et al. [24], we let $\tau_k = \frac{k}{K+1}$, $k = 1,2,\ldots,K$ for some K. By weighting the different loss functions in CQR with the IPW method, our first WCQR estimator is defined as

$$(\hat{\mathbf{b}},\hat{\beta}_{\mathrm{WCQR1}}) = \mathrm{argmin}_{b,\beta}L_n(\hat{\pi}(U),\beta,\mathbf{b}), \tag{2.6}$$

where $L_n(\hat{\pi}(U),\beta,\mathbf{b}) = \sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k})$, the weight ω_k's are allowed to be negative and satisfy $\sum_{k=1}^{K}\omega_k g(b_{\tau_k}) > 0$, where $g(\cdot)$ is the density function of the error term ε.

The following theorem presents the asymptotic distribution of $\hat{\beta}_{\text{WCQR1}}$. We first introduce some notations. Let β^* be the true value of β, $b^*_{\tau_k}$ be the τ_k-th quantile of ε and $\mathbf{b}^* = (b^*_{\tau_1}, b^*_{\tau_2}, \ldots, b^*_{\tau_K})^T$. Denote $f_i^* = f(X_i,\beta^*)$, $\nabla f_i^* = \frac{\partial f(X_i,\beta)}{\partial \beta}|_{\beta=\beta^*}$, $\Sigma_1 = E[\nabla f_1^*(\nabla f_1^*)^T]$, $\Sigma_2 = E[\frac{\nabla f_1^*(\nabla f_1^*)^T}{\pi(U)}]$, $\mathbf{g} = (g(b^*_{\tau_1}), g(b^*_{\tau_2}), \ldots, g(b^*_{\tau_K}))^T$, $\mathbf{\Omega} = \{\min(\tau_k, \tau_{k'})(1-\max(\tau_k, \tau_{k'}))\}_{1\le k,k'\le K}$, and $\mathbf{H} = (\frac{\min(\tau_k,\tau_{k'})(1-\max(\tau_k,\tau_{k'}))}{g(b^*_{\tau_k})g(b^*_{\tau_{k'}})})_{1\le k,k'\le K}$.

Theorem 2.1. *Suppose that the conditions $C1-C6$ in Appendix hold and β^* is the true value. Then we have*

$$\sqrt{n}(\hat{\beta}_{WCQR1} - \beta^*) \xrightarrow{D} N\left(0, \frac{\omega^T\Omega\omega}{\omega^T\mathbf{g}\mathbf{g}^T\omega}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right).$$

Similar to Jiang et al. [24] and Zhao et al. [26], we can derive the optimal weights by minimizing $\frac{\omega^T\Omega\omega}{\omega^T\mathbf{g}\mathbf{g}^T\omega}$ in the asymptotic variance given in Theorem 2.1.

Corollary 2.1. *The optimal weight vector $\omega^* = (\omega_1^*, \omega_2^*, \cdots, \omega_K^*)^T$ for $\hat{\beta}_{WCQR1}$ is*

$$\omega^* = argmin\frac{\omega^T\Omega\omega}{\omega^T\mathbf{g}\mathbf{g}^T\omega} = (\mathbf{g}^T\Omega^{-2}\mathbf{g})^{-1/2}\Omega^{-1}\mathbf{g}. \tag{2.7}$$

Note that the optimal weight depends on the density function of ε. Based on estimated residuals $\hat{\varepsilon}_i$, the usual nonparametric density estimation methods can provide a consistent estimator $\hat{g}(\cdot)$ of $g(\cdot)$. Then the estimated optimal weight vector is $\hat{\omega}^* = (\hat{\mathbf{g}}^T\Omega^{-2}\hat{\mathbf{g}})^{-1/2}\Omega^{-1}\hat{\mathbf{g}}$. With the optimal weight vector $\hat{\omega}^*$ obtained in hand, the first optimal WCQR estimator of β is defined as

$$\hat{\beta}_{\text{OWCQ1}} = \arg\min_{\beta}\sum_{k=1}^{K}\hat{\omega}_k^*\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - \hat{b}_{\tau_k}). \tag{2.8}$$

Corollary 2.2. *The optimal weighted compositive quantile estimators $\hat{\beta}_{OWCQ1}$ of β has the optimal asymptotic variance $\frac{1}{n}(\mathbf{g}^T\Omega^{-1}\mathbf{g})^{-1}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$.*

Next, we present the second weighting schemes. Our method is inspired by Wang, Song and Zhang [23]. Let

$$(\hat{b}_{\tau_k}, \hat{\beta}_{\tau_k}) = \arg\min_{b_{\tau_k},\beta}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k}),$$

then the second WCQR estimator is defined as

$$\hat{\beta}_{\text{WCQR2}} = \sum_{k=1}^{K}\omega_k\hat{\beta}_{\tau_k}, \tag{2.9}$$

where ω_k's satisfy $\sum_{k=1}^{K}\omega_k = 1$. The asymptotic distribution of $\hat{\beta}_{\text{WCQR2}}$ is summarized in the following theorem.

Theorem 2.2. *Suppose that the conditions $C1-C6$ in Appendix hold and β^* be is true parameter value. Then we have*

$$\sqrt{n}(\hat{\beta}_{WCQR2} - \beta^*) \xrightarrow{\mathcal{D}} N\left(0, \omega^T\boldsymbol{H}\omega\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right).$$

Similarly, we can obtain the optimal weight by minimizing $\omega^T\mathbf{H}\omega$ in the asymptotic covariance given in Theorem 2.2. As a result, the second optimal WCQR of β can be correspondingly defined as $\hat{\beta}_{\text{OWCQ2}}$ with the associated optimal asymptotic variance derived in the following corollary.

Corollary 2.3. *The optimal weight vector $\omega^* = (\omega_1^*, \omega_2^*, \ldots, \omega_K^*)^T$ of WCQR2 is*

$$\omega^* = \arg\min_{\omega^T 1 = 1} \omega^T \boldsymbol{H} \omega = \frac{\boldsymbol{H}^{-1}\boldsymbol{1}}{\boldsymbol{1}^T \boldsymbol{H}^{-1}\boldsymbol{1}}, \tag{2.10}$$

where $\boldsymbol{1}$ is a $K \times 1$ vector with all elements 1. With this optimal weight vector, the optimal WCQR estimator $\hat{\beta}_{OWCQ2} = \sum_{k=1}^{K} \omega_k^* \hat{\beta}_{\tau_k}$ has the optimal asymptotic variance

$$\frac{1}{n}(\boldsymbol{1}^T \boldsymbol{H}^{-1}\boldsymbol{1})^{-1}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} = \frac{1}{n}(\mathbf{g}^T \Omega^{-1}\mathbf{g})^{-1}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}.$$

Remark 1. The optimal weight of OWCQ2 is essentially the same as Zhao and Lian [25], but with different representation. And from the above results for the two weighting methods we observe that if we use the optimal weight vectors, the optimal WCQR estimators achieve the same optimal asymptotic variance $\frac{1}{n}(\mathbf{g}^T \Omega^{-1}\mathbf{g})^{-1}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$.

3. Simulation studies

In this section, we use simulation studies to examine the finite sample performance of our proposed methods and compare it with the inverse probability weighted CQR (IWCQ) method which uses the same weight for different QR models, and the inverse probability WLS estimator. Referring to Zou and Yuan [6], the estimator of the proposed methodology is nearly efficient as the oracle maximum likelihood (OML) estimator for $K \geq 9$ in various error distributions. Therefore, we take $K = 10$, $\tau_k = k/11$, $k = 1, 2, \ldots, 10$, and consider the exponential regression models

$$Y = \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) + \varepsilon,$$

where $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1$ and (X_1, X_2, X_3) follows multivariate normal distribution with covariances always 0.5 and variances always 1. The model error ε and $X = (X_1, X_2, X_3)^T$ are independent. Then, using the method described in Section 4 in Wang, Chen and Lin [27], we set the data in X_3 to be missing at random while X_1, X_2, Y are fully observed. And we consider two selection probability functions

$$\pi_1(X_1, X_2) = \exp(2 + 0.5X_1 + 0.5X_2)/\left[1 + \exp(2 + 0.5X_1 + 0.5X_2)\right],$$
$$\pi_2(X_1, X_2) = \exp(1 + 1.25X_1 + X_2)/\left[1 + \exp(1 + 1.25X_1 + X_2)\right].$$

Their corresponding average missing rates are 15% and 35% respectively. In our simulation, four different distributions of model error ε are considered:

(Case 1) The standard normal distribution $N(0, 1)$.

(Case 2) The centralized t distribution with four degrees of freedom.

(Case 3) The mixture of normal distribution $0.6N(0, 1) + 0.4N(2, 1)$.

(Case 4) The centralized χ^2 distribution with four degrees of freedom.

In the simulation, samples of size $n = 200$ and $n = 600$ are generated independently. Four estimation methods, OWCQ1, OWCQ2, WLS and IWCQ are used to estimate β_1, β_2 and β_3 under the above selection probability functions and error distributions. Then the root of mean squared errors

(RMSEs) can be calculated. To evaluate the different estimators, we repeat the process 1000 times and calculate the average RMSEs. The simulation results are reported in Tables 1–4 for cases that the selection probability function $\pi(\cdot)$ is known (denoted as T), estimated nonparametrically (denoted as N) and parametrically (denoted as P). When the selection probability is estimated nonparametrically, we use the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$ to construct the multiplicative kernel $L(x_1, x_2) = K(x_1)K(x_2)$, and use the bandwidth proposed by Ruppert, Sheather and Wand [28]. When $\pi(\cdot)$ is estimated by parametric method, we apply model (2.5) to estimate it. Meanwhile, similar to Jiang et al. [24], our proposed estimator involves a weighting scheme and the density of error is known in simulations, so we took the optimal weight ω^* (see Section 2.2) for all simulations.

Table 1. The RMSEs (multiplied by 10^4) for β under the selection probability function $\pi_1(X_1; X_2)$ for $n = 200$.

ε	β	OWCQ1			OWCQ2			WLS			IWCQ		
		T	N	P	T	N	P	T	N	P	T	N	P
Case1	β_1	69.815	70.083	70.394	72.635	72.568	72.357	69.481	69.228	69.405	70.208	69.758	70.045
	β_2	67.619	67.682	67.697	70.266	70.334	70.316	66.842	66.872	66.880	67.598	67.771	67.520
	β_3	67.451	66.999	67.862	69.436	69.148	69.571	66.548	66.339	66.534	67.456	67.096	67.599
Case2	β_1	91.124	90.406	91.482	92.599	92.082	92.282	97.444	97.329	97.498	91.918	91.159	91.469
	β_2	86.790	84.551	86.247	85.497	84.982	85.157	93.340	92.772	93.262	86.649	86.011	86.640
	β_3	87.330	86.256	86.832	86.259	86.408	86.438	92.537	92.775	92.519	87.302	87.413	86.463
Case3	β_1	103.49	103.60	104.43	104.01	104.21	104.02	116.86	116.63	116.76	103.24	104.69	104.19
	β_2	97.220	96.335	97.271	94.204	94.251	94.005	112.09	111.73	112.13	99.724	100.34	99.991
	β_3	103.11	103.66	104.10	100.51	100.17	100.26	117.34	117.38	117.33	105.15	103.92	105.87
Case4	β_1	194.52	192.19	195.21	168.33	170.55	171.35	411.30	411.00	409.41	256.99	259.41	259.73
	β_2	178.00	168.97	174.80	150.73	148.00	149.04	391.41	390.18	391.30	241.40	238.28	241.31
	β_3	196.63	195.57	196.35	163.73	167.43	169.52	400.91	400.87	399.81	250.51	247.66	247.75

Table 2. The RMSEs (multiplied by 10^4) for β under the selection probability function $\pi_2(X_1; X_2)$ for $n = 200$.

ε	β	OWCQ1			OWCQ2			WLS			IWCQ		
		T	N	P	T	N	P	T	N	P	T	N	P
Case1	β_1	72.059	70.502	72.837	72.631	72.315	72.331	68.665	68.472	68.565	73.059	71.171	72.846
	β_2	69.884	66.711	69.543	69.375	69.399	69.523	65.568	65.516	65.486	69.057	67.296	68.708
	β_3	70.649	69.247	71.376	71.898	70.976	71.581	67.101	66.943	67.111	70.199	68.972	70.806
Case2	β_1	89.885	90.720	91.987	91.404	90.922	89.848	95.801	95.740	95.810	94.513	91.495	94.635
	β_2	84.532	83.575	85.351	81.752	81.301	82.248	89.878	89.691	89.885	84.997	81.760	84.635
	β_3	86.637	86.019	88.037	86.442	85.448	85.523	90.878	91.106	90.837	88.500	85.344	88.625
Case3	β_1	111.25	107.40	110.85	106.60	105.38	106.70	117.16	117.54	117.09	114.25	110.63	110.23
	β_2	102.81	98.443	103.60	95.112	96.201	95.374	111.66	111.44	111.53	108.05	102.58	107.68
	β_3	109.44	106.47	105.94	100.80	101.37	100.26	115.30	115.45	115.55	113.12	106.08	108.44
Case4	β_1	200.03	190.67	196.16	178.68	173.67	185.99	410.08	412.02	410.44	279.83	264.94	285.81
	β_2	170.64	167.98	174.57	155.36	146.81	153.24	382.39	381.07	382.02	259.52	239.58	250.77
	β_3	196.65	186.94	195.98	175.10	167.23	175.20	391.13	391.01	391.34	270.68	257.52	270.36

Table 3. The RMSEs (multiplied by 10^4) for β under the selection probability function $\pi_1(X_1; X_2)$ for $n = 600$.

ε	β	OWCQ1			OWCQ2			WLS			IWCQ		
		T	N	P	T	N	P	T	N	P	T	N	P
Case1	β_1	25.572	25.515	24.947	26.134	26.437	26.173	24.447	24.525	24.442	25.424	25.342	25.320
	β_2	23.663	23.839	23.743	24.582	24.823	24.726	23.140	23.253	23.142	23.938	23.643	23.878
	β_3	24.046	23.974	23.934	24.758	24.746	24.892	23.426	23.507	23.429	24.246	24.019	24.215
Case2	β_1	30.219	30.031	30.487	30.349	30.274	30.454	32.960	32.914	32.959	30.739	30.452	30.924
	β_2	30.093	30.356	30.087	29.949	29.838	29.894	32.766	32.736	32.772	30.636	30.554	30.107
	β_3	29.388	29.358	29.372	29.042	29.074	29.069	32.962	32.924	32.966	29.949	29.783	29.989
Case3	β_1	41.449	41.209	41.214	37.899	37.862	37.805	48.247	48.244	48.183	41.469	41.915	42.827
	β_2	37.420	37.788	37.515	35.155	34.787	34.723	46.331	46.481	46.301	39.647	38.938	40.156
	β_3	37.194	36.275	36.613	33.889	33.677	33.792	45.890	45.957	45.905	37.930	37.878	38.686
Case4	β_1	69.441	67.445	68.620	56.757	56.679	56.131	173.27	173.52	173.24	104.62	106.89	104.78
	β_2	65.303	63.240	62.968	52.234	50.795	51.504	176.16	176.44	176.46	105.95	107.51	103.95
	β_3	67.368	62.231	62.897	54.403	54.812	54.689	182.13	182.04	182.06	104.29	111.02	103.67

Table 4. The RMSEs (multiplied by 10^4) for β under the selection probability function $\pi_2(X_1; X_2)$ for $n = 600$.

ε	β	OWCQ1			OWCQ2			WLS			IWCQ		
		T	N	P	T	N	P	T	N	P	T	N	P
Case1	β_1	26.185	25.492	26.202	25.716	25.794	25.532	24.259	24.271	24.288	26.034	25.786	25.998
	β_2	24.210	24.086	24.215	24.426	24.648	24.535	22.966	23.048	22.968	24.498	24.623	24.252
	β_3	24.971	24.826	25.319	25.077	25.128	25.043	23.276	23.295	23.311	25.454	24.999	25.898
Case2	β_1	31.548	31.632	31.631	30.495	30.447	30.524	33.423	33.397	33.411	32.692	32.331	32.422
	β_2	30.301	29.961	30.383	29.742	29.760	29.827	32.471	32.430	32.454	31.721	31.218	31.278
	β_3	30.325	29.495	29.733	29.124	28.802	29.064	32.906	32.873	32.897	31.516	30.954	30.065
Case3	β_1	45.130	43.852	45.128	37.870	37.555	38.020	48.043	48.115	48.048	50.746	47.710	52.185
	β_2	41.167	39.934	41.160	35.190	35.213	35.411	46.267	46.312	46.233	45.117	43.963	47.649
	β_3	40.974	38.909	40.893	34.035	34.065	33.732	45.492	45.414	45.479	44.275	43.626	45.085
Case4	β_1	78.214	75.849	77.303	57.448	57.145	56.259	171.73	172.81	171.66	134.52	124.10	136.22
	β_2	73.627	69.376	70.722	52.432	51.211	51.105	175.48	175.61	175.55	136.08	124.71	136.97
	β_3	73.030	69.922	69.207	54.729	52.922	54.132	179.53	179.19	179.42	134.62	125.02	135.85

From Tables 1–4 we observe that when the model error ε follows the standard normal distribution $N(0, 1)$, WLS performs the best among the four estimators considered, while OWCQ1, OWCQ2 and IWCQ behave very similarly. For all other non-normal distributions considered, WLS always performs the worst. The performance of the other three methods are very similar when the model error follows the centralized t distribution with four degrees of freedom. It is further noted that when the missing rate is high or the sample size is large, our proposed methods are superior to IWCQ. Particularly, when the model error follows chi-square distribution with four degrees of freedom, the superiority of both OWCQ1 and OWCQ2 are even more obvious. We also find that for OWCQ1 and IWCQ methods a better result can be obtained by estimating the selection probability function with a nonparametric method. At the same time, IWCQ also performs much better than WLS.

When sample size is large, it can be seen from Tables 3 and 4 that the performance of the four estimators are significantly improved compared with that when the sample case is small.And our proposed estimators have more obvious advantages over WLS and IWCQ. We observe that both OWCQ1 and OWCQ2 always have a high accuracy under any of the four error distributions, and OWCQ2 performs slightly better than OWCQ1 except when the model error ε follows the standard normal distribution. We also find that the RMSEs are not sensitive to missing rate. In addition, the calculation speed of OWCQ1 is faster than OWCQ2 when the optimal weight obtained from the known error distribution is used. For example, when we simulated case1 at n=200 and π=0.15, we found that OWCQ1 was about 20% faster than OWCQ2. For other cases, the difference between OWCQ1 and OWCQ2 in computing speed is similar.

4. A real data example

In this section, we will illustrate our proposed methods using a real data originally presented by Baum [29] to investigate how age, marriage state, number of children and education background affect whether a women works or not. For each women there are five variables:

- Work (y): 1=Yes, 0=Not;
- Age (x_1): the age of the women;
- Children(x_2): the number of the children the women raises;
- Education (x_3): the years that the women has passed in school;
- Married (x_4): 1=Yes, 0=Not.

Note that the response y is the average estimated probability of work. A logistic model with all of covariates given by

$$y_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i})} + \varepsilon_i, \quad i = 1, 2, \ldots, 2000$$

is suitable for modeling the relationship between the choice of work and all possible factors. In order to use the data set to illustrate our methods, artificial missing data were created by using the selection probability $\pi(X) = \frac{\exp(\gamma_0+\gamma_1 x_1+\gamma_2 x_2)}{1+\exp(\gamma_0+\gamma_1 x_1+\gamma_2 x_2)}$. The missing proportion is about 18.65% with $\gamma_0 = 2, \gamma_1 = 0.15, \gamma_2 = 0.25$, and, following Li and Ding [30], the quantile vector is taken as $\tau = (0.2, 0.4, 0.6, 0.8)^T$ with $K = 4$.

From (2.7) and (2.10), we know that the optimal weights depend on $g(b^*_\tau)$ and b^*_τ, both of which are unknown here and need to be estimated. Motivated by Sun and Sun [31] and Zhao and Xiao [7], we propose the following procedure under the case when the selection probability is known.

(1) Use the uniform weight $\omega = (1/K, 1/K, \ldots, 1/K)^T$ to obtain the preliminary estimator $\hat{\beta}$ of β as follows:

$$(\hat{b}_{\tau_1}, \hat{b}_{\tau_1}, \ldots, \hat{b}_{\tau_K}, \hat{\beta}) = \arg\min_{b_{\tau_k},\beta} \sum_{k=1}^{K} \frac{1}{K} \sum_{i=1}^{n} \frac{\delta_i}{\pi(U_i)} \rho_{\tau_k}(Y_i - b_{\tau_k} - f(X_i, \beta)).$$

(2) Let $m = \sum_{i=1}^{n} \delta_i$. Without loss of generality, we assume the first m observations are complete. Then, based on the complete data, the pseudo residuals $\hat{\varepsilon}_i$ with $\delta_i = 1$ are computed as $\hat{\varepsilon}_i = \frac{\delta_i}{\pi(U_i)}(Y_i - f(X_i, \hat{\beta}))$, $i = 1, 2, \ldots, m$.

(3) Use the nonparametric kernel density estimator to estimate $g(t)$:

$$\hat{g}(t) = \frac{1}{mb} \sum_{i=1}^{m} K(\frac{t - \hat{\varepsilon}_i}{b}),$$

where $K(\cdot)$ is a non-negative kernel function and the bandwidth b is selected by

$$b = 0.9 \times \min\left\{ \mathrm{SD}(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_m), \frac{\mathrm{IQR}(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_m)}{1.34} \right\} \times m^{-1/5},$$

where SD and IQR stand for the sample standard deviation and sample interquantile range, respectively.

(4) Estimate $g(b^*_{\tau_k})$ by $\hat{g}(\hat{b}_{\tau_k})$ and then substitute it into (2.7) or (2.10), from which the optimal weight vector can be obtained, where $\hat{b}_{\tau_k}$ denotes the sample τ_k-quantile of $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_m$.

It is obvious that when a women has a work, the response y_i will take a larger value. Because there are only 32.85% of women in the data set does not work, we could believe that a woman has a job if the corresponding response $\hat{y}_i$ is bigger than the 0.3285 quantile of the fitted values $\hat{y}$. In order to compare the performance of our proposed methods with IWCQ and the composite quantile estimator which only uses the fully observed data (denoted by CQR-CCA), we calculate the fitted values $\hat{y}$ with all the 2000 data of the above four methods respectively, and predict whether a women works or not. The prediction accuracy is reported in Table 5. From Table 5 we observe that IWCQ method can obviously improve the efficiency of estimation in the case of missing data, and CQR-CCA estimator has the lowest accuracy. It is obvious that our proposed methods are more accurate compared with IWCQ method.

Table 5. Accuracy of prediction.

	OWCQ1	OWCQ2	IWCQ	CQR-CCA
Accuracy	0.708	0.693	0.6725	0.6195

5. Discussion

In this article, we have proposed two types of weighted quantile estimators for nonlinear models with missing covariates. The asymptotic properties of our proposals have been obtained under certain conditions. Our simulation studies reveal that our proposed method has better advantages than the existing methods. Finally, we propose some future directions. First, We only consider the estimates of unknown parameters in this paper, and future studies can start from variable selection. Second, the logistic model for the selection probability function is assumed in our article. When the selection probability function is misspecified, how to derive a robust estimation of the selection probability could be a direction for further study. Third, our method could be used in Altun et al. [32] to obtain the unknown model parameters of new extended gamma distribution. At last, how to generalize our method to optimal reinsurance problems of Fang, Cheng and Qu [33] is also an interesting topic.

Acknowledgments

The research is supported by NSF projects (ZR2021MA077, ZR2021MA048 and ZR2019MA016) of Shandong Province of China.

Conflict of interest

All authors declare that there is no conflict of interest.

References

1. D. L. Wang, H. L. Xu, Q. Wu, Averaging versus voting: A comparative study of strategies for distributed classification, *Math. Found. Comput.*, **3** (2020), 185–193. http://dx.doi.org/10.3934/mfc.2020017
2. W. Zhao, H. Lian, S. Ma, Robust reduced-rank modeling via rank regression, *J. Stat. Plan. Infer.*, **180** (2017), 1–12. http://dx.doi.org/10.1016/j.jspi.2016.08.009
3. F. Zhang, R. Li, H. Lian, D. Bandyopadhyay, Sparse reduced-rank regression for multivariate varying-coefficient models, *J. Stat. Comput. Simul.*, **91** (2021), 752–767. http://dx.doi.org/10.1080/00949655.2020.1829622
4. T. L. Gong, C. Xu, H. Chen, Modal additive models with data-driven structure identification., *Math. Found. Comput.*, **3** (2020), 165–183. http://dx.doi.org/10.3934/mfc.2020016
5. R. Koenker, G. W. Bassett, Regression quantiles, *Econometrica*, **46** (1978), 33–50. http://dx.doi.org/10.2307/1913643
6. H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, *Ann. Stat.*, **36** (2008), 1108–1126. http://dx.doi.org/10.1214/07-AOS507
7. Z. Zhao, Z. Xiao, Efficient regressions via optimally combining quantile information, *Economet. Theory*, **30** (2014), 1272–1314. http://dx.doi.org/10.1017/S0266466614000176
8. R. Koenker, A note on L-estimates for linear models, *Statist. Probab. Lett.*, **2** (1984), 323–325. http://dx.doi.org/10.1016/0167-7152(84)90040-3
9. R. Koenker, *Quantile regression*, Cambridge: Cambridge University Press, 2005. http://dx.doi.org/10.1017/CBO9780511754098
10. X. J. Jiang, J. Jiang, X. Song, Oracle model selection for nonlinear models based on weighted composite quantile regression, *Stat. Sin.*, **22** (2012), 1479–1506. http://dx.doi.org/10.5705/ss.2010.203
11. D. Bloznelis, G. Claeskens, J. Zhou, Composite versus model-averaged quantile regression, *J. Stat. Plan. Infer.*, **200** (2019), 32–46. http://dx.doi.org/10.1016/j.jspi.2018.09.003
12. F. Yates, The analysis of replicated experiments when the field results are incomplete, *Emprie Jour. Exp. Agric.*, **1** (1933), 129–142.
13. L. Q. Xia, X. L. Wang, P. X. Zhao, Y. Q. Song, Empirical likelihood for varying coefficient partially nonlinear model with missing responses, *AIMS Mathematics*, **6** (2021), 7125–7152. http://dx.doi.org/10.3934/math.2021418

14. D. G. Horvitz, D. J. Thompson, A generalization of sampling without replacement from a finite universe, *J. Am. Stat. Assoc.*, **47** (1952), 663–685. http://dx.doi.org/10.1080/01621459.1952.10483446

15. D. B. Rubin, Inference and missing data, *Biometrika*, **63** (1976), 581–592. http://dx.doi.org/10.1093/biomet/63.3.581

16. R. J. A. Little, D. B. Rubin, *Statistical analysis with missing data*, 2 Eds., New York: Wiley, 2002. http://dx.doi.org/10.1002/9781119013563

17. J. M. Robins, A. Rotnitzky, L. P. Zhao, Estimation of regression coefficients when some of regression coefficients estimation regressors are not always observed, *J. Am. Stat. Assoc.*, **89** (1994), 846–866. http://dx.doi.org/10.2307/2290910

18. J. G. Ibrahim, H. T. Zhu, N. S. Tang, Model selection criteria for missing data problems via the EM algorithm, *J. Am. Stat. Assoc.*, **103** (2008), 1648–1658. http://dx.doi.org/10.1198/016214508000001057

19. J. Qin, J. Shao, B. Zhang, Efficient and doubly robust imputation for covariate-dependent missing responses, *J. Am. Stat. Assoc.*, **103** (2008), 797–810. http://dx.doi.org/10.1198/016214508000000238

20. B. Sherwood, L. Wang, X. H. Zhou, Weighted quantile regression for analyzing health care cost data with missing covariates, *Stat. Med.*, **32** (2013), 4967–4979. http://dx.doi.org/10.1002/sim.5883

21. X. R. Chen, A. T. Wan, Y. Zhou, Efficient quantile regression analysis with missing observations, *J. Am. Stat. Assoc.*, **110** (2015), 723–741. http://dx.doi.org/10.1080/01621459.2014.928219

22. H. Yang, H. L. Liu, Penalized weighted composite quantile estimators with missing covariates, *Stat. Papers*, **57** (2014), 69–88. http://dx.doi.org/10.1007/s00362-014-0642-2

23. X. L. Wang, Y. Q. Song, S. X. Zhang, An efficient estimation for the parameter in additive partially linear models with missing covariates, *J. Korean Stat. Soc.*, **49** (2020), 779–801. http://dx.doi.org/10.1007/s42952-019-00036-6

24. X. J. Jiang, J. Z. Li, T. Xia, W. F. Yan, Robust and efficient estimation with weighted composite quantile regression, *Physica A*, **457** (2016), 413–423. http://dx.doi.org/10.1016/j.physa.2016.03.056

25. K. Zhao, H. Lian, A note on the efficiency of composite quantile regression, *J. Stat. Comput. Simul.*, **86** (2016), 1334–1341. http://dx.doi.org/10.1080/00949655.2015.1062096

26. W. Zhao, H. Lian, M. Chen, X. Song, Composite quantile regression for correlated data, *Comput. Stat. Data Anal.*, **109** (2009), 15–33. http://dx.doi.org/10.1016/j.csda.2016.11.015

27. X. L. Wang, F. Chen, L. Lin, Empirical likelihood inference for estimating equation with missing data, *Sci. China. Math.*, **56** (2013), 1233–1245. http://dx.doi.org/10.1007/s11425-012-4504-x

28. D. Ruppert, S. J. Sheather, M. P. Wand, An effective bandwidth selector for local least squares regression, *J. Am. Stat. Assoc.*, **90** (1995), 1257–1270. http://dx.doi.org/10.1080/01621459.1995.10476630

29. C. F. Baum, *An introduction to modern econometrics using Stata*, Texas: Stata Press, 2006.

30. Y. Li, J. Ding, Weighted composite quantile regression method via empirical likelihood for non linear models, *Commun. Stat.-Theor. M.*, **47** (2018), 4286–4296. http://dx.doi.org/10.1080/03610926.2017.1373816

31. J. Sun, Q. H. Sun, An improved and efficient estimation method for varying-coefficient model with missing covariates, *Statist. Probab. Lett.*, **105** (2015), 296–303. http://dx.doi.org/10.1016/j.spl.2015.09.009

32. E. Altun, M. Korkmaz, M. Elmorshedy, M. S. Eliwa, The extended gamma distribution with regression model and applications, *AIMS Mathematics*, **6** (2021), 2418–2439. http://dx.doi.org/10.3934/math.2021147

33. Y. Fang, G. Cheng, Z. F. Qu, Optimal reinsurance for both an insurer and a reinsurer under general premium principles, *AIMS Mathematics*, **5** (2020), 3231–3255. http://dx.doi.org/10.3934/math.2020208

34. K. Knight, Limiting distributions for L1 regression estimators under general conditions, *Ann. Statist.*, **26** (1998), 755–770. http://dx.doi.org/10.1214/aos/1028144858

35. H. Wong, S. Guo, M. Chen, W. C. Ip, On locally weighted estimation and hypothesis testing of varyingcoefficient models with missing covariates, *J. Stat. Plan. Infer.*, **139** (2009), 2933–2951. http://dx.doi.org/10.1016/j.jspi.2009.01.016

A. Appendix: Assumptions and proofs

The following assumptions are needed in the proof of our main results, and they are commonly used in nonparametric regression and quantile regression literatures.

C1: The random error ε is independent of X, its density function $g(\cdot)$ has a continuous and uniformly bounded derivative. For each p-vector $\mathbf{u}$, the cumulative distribution function $G(\cdot)$ satisfies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{u_0+(\nabla f_i^*)^T\mathbf{u}}\sqrt{n}\left[G(a+\frac{t}{\sqrt{n}})-G(a)\right]dt=\frac{1}{2}g(a)(u_0,\mathbf{u}^T)\begin{pmatrix}1 & 0\\ 0 & \Sigma_1\end{pmatrix}(u_0,\mathbf{u}^T)^T.$$

C2: There is a large enough open subset $\mathcal{B}$ contains the true parameter point β_0, such that for all X_i the second derivative matrix $\nabla^2 f(X_i,\beta)$ of $f(X_i,\beta)$ with respect to β satisfies that

$$||\nabla^2 f(X_i,\beta_1)-\nabla^2 f(X_i,\beta_2)||\leq M(X_i)||\beta_1-\beta_2||$$

and $|\partial^2 f(X_i,\beta)/\partial\beta_j\partial\beta_k|\leq N_{ij}(X_i)$ for all $\beta\in\mathcal{B}$, where $E[M^2(X_i)]<\infty$ and $E[N_{ij}^2(X_i)]<C_1<\infty$ for all j,k.

C3: The matrix Σ_1 is positive definite.

C4: The function $\pi(\cdot)$ has bounded joint derivatives up to order r_0 ($r_0>q$) and satisfies $\min_i\pi(U_i)\geq c_0$ for some $c_0>0$.

C5: The density function $P_U(\cdot)$ of U has bounded joint derivatives up to order r_0 ($r_0>q$) on the support of U and satisfies $P_U(u)\geq c$ for some $c>0$.

C6: The kernel functions $K(\cdot)$ is a bounded kernel function of order r_0 with bounded support. The bandwidth h satisfies $h\to 0$, $nh^{2r_0}\to 0$, $nh^{2q}/(\log n\cdot\log\log n)\to\infty$ as $n\to\infty$.

C7: The MLE $\hat{\gamma}$ of γ is $\sqrt{n}$-consistent and satisfies the regularity conditions of asymptotic normality.

Assumptions C1 and C2 are the same as the assumptions (b)and (c) in Jiang et al. [24]. Assumptions C1, C4, C5 and C6 are commonly used conditions in missing data and nonparametric regression.

Proof of Theorem 2.1. The proof of Theorem 2.1 is divided into three parts: (1) when the selection probability function is supposed to be known, (2) when the selection probability function is estimated by the nonparametric method, and (3) when the selection probability function is estimated by the parametric method.

(1) Let $\sqrt{n}(\hat{\beta}-\beta^*) = \mathbf{u}$, $\sqrt{n}(\hat{b}_{\tau_k}-b_{\tau_k}) = v_k$, $\mathbf{v} = (v_1, v_2, \ldots, v_K)^T$, $\theta = (\mathbf{u}^T, \mathbf{v}^T)$, $S_n(\pi(U), \theta) = L_n(\pi(U), \beta^* + n^{-1/2}\mathbf{u}, \mathbf{b}^* + n^{-1/2}\mathbf{v}) - L_n(\pi(U), \beta^*, \mathbf{b}^*)$, $\xi_i(\mathbf{u}, v_k) = (\nabla f_i^*)^T\mathbf{u} + v_k$, $\eta_i = \sum_{k=1}^K \omega_k[I(\varepsilon_i < b^*_{\tau_k}) - \tau_k]$. Then minimizing the object function $L_n(\pi(U), \beta, \mathbf{b})$ is equivalent to minimizing $S_n(\pi(U), \theta)$. Similar to Jiang et al. [24], we let

$$\begin{aligned} S_n^*(\pi(U), \theta) &= \sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)}\{\rho_{\tau_k}(Y_i - f_i^* - (\nabla f_i^*)^T n^{-1/2}\mathbf{u} - (b^*_{\tau_k} + n^{-1/2}v_k)) \\ &\quad -\rho_{\tau_k}(Y_i - f_i^* - b^*_{\tau_k})\}, \end{aligned}$$

$$\begin{aligned} S_n^{**}(\pi(U), \theta) &= \sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)}\{\rho_{\tau_k}(Y_i - f_i^* - (\nabla f_i^*)^T n^{-1/2}\mathbf{u} - \frac{1}{2n}\mathbf{u}^T(\nabla^2 f_i^*)\mathbf{u} \\ &\quad -(b^*_{\tau_k} + n^{-1/2}v_k)) - \rho_{\tau_k}(Y_i - f_i^* - b^*_{\tau_k})\}. \end{aligned}$$

Then

$$S_n^*(\pi(U), \theta) = \sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)}\{\rho_{\tau_k}(\varepsilon_i - b^*_{\tau_k} - \frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)) - \rho_{\tau_k}(\varepsilon_i - b^*_{\tau_k})\},$$

$$S_n^{**}(\pi(U), \theta) = \textstyle\sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)}\{\rho_{\tau_k}(\varepsilon_i - b^*_{\tau_k} - \frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k) - \frac{1}{2n}\mathbf{u}^T(\nabla^2 f_i^*)\mathbf{u}) - \rho_{\tau_k}(\varepsilon_i - b^*_{\tau_k})\}.$$

Applying the identity in Knight [34]

$$\rho_\tau(r - s) - \rho_\tau(r) = s(I(r < 0) - \tau) + \int_0^s [I(r \le t) - I(r \le 0)]dt,$$

we have

$$\begin{aligned} S_n^*(\pi(U), \theta) &= \sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)}\{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)[I(\varepsilon_i < b^*_{\tau_k}) - \tau_k]\} \\ &\quad +\sum_{k=1}^K \omega_k \sum_{i=1}^n \frac{\delta_i}{\pi(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})]dx \end{aligned}$$

$$= \sum_{k=1}^{K} \zeta_{nk} v_k + Z_n^T \mathbf{u} + \sum_{k=1}^{K} \omega_k B_{nk},$$

where

$$\zeta_{nk} = \omega_k \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi(U_i)} [I(\varepsilon_i < b^*_{\tau_k}) - \tau_k],$$

$$Z_n = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} \tfrac{\delta_i}{\pi(U_i)} (\nabla f_i^*)^T \eta_i,$$

$$B_{nk} = \textstyle\sum_{i=1}^{n} \tfrac{\delta_i}{\pi(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})]dx.$$

Let

$$B_{ni,k} = \frac{\delta_i}{\pi(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})]dx,$$

for any $\epsilon > 0$, we have

$$[B_{ni,k}]^2 = [B_{ni,k}]^2 I(\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k) \geq \epsilon) + [B_{ni,k}]^2 I(\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k) < \epsilon).$$

Similar to the proof of Theorem 2.1 in Yang and Liu [22], we can verify that $\sum_{i=1}^{n}[B_{ni,k} - E(B_{ni,k})] = o_p(1)$. Moreover, we have

$$\begin{aligned} E(B_{nk}) &= \sum_{i=1}^{n} E(B_{ni,k}) = nE\{\frac{\delta_i}{\pi(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})]dx\} \\ &= nE\{\int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)} [G(b^*_{\tau_k} + x) - G(b^*_{\tau_k})]dx\} \\ &= \frac{1}{2} g(b^*_{\tau_k})[v_k^2 + \mathbf{u}^T \Sigma_1 \mathbf{u}] + o_p(1). \end{aligned}$$

Note that $E\{\frac{\delta_i}{\pi(U_i)}[I(\varepsilon_i < b^*_{\tau_k}) - \tau_k]\} = 0$, using the central limit theorem, we know ζ_{nk} and Z_n converge in distribution to ζ_k and Z, where ζ_k is normal with mean 0 and $Z \sim N(\mathbf{0}, \Sigma_2(\omega^T \Omega \omega))$. So we have

$$S_n^*(\pi(U), \theta) \to^d S(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{K} \zeta_k v_k + Z^T \mathbf{u} + \frac{1}{2} \sum_{k=1}^{K} \omega_k g(b^*_{\tau_k})[v_k^2 + \mathbf{u}^T \Sigma_1 \mathbf{u}].$$

Similar to Jiang et al. [24], we can prove that

$$S_n^{**}(\pi(U), \theta) - S_n^*(\pi(U), \theta) = o_p(1)$$

and

$$S_n(\pi(U), \theta) - S_n^*(\pi(U), \theta) = o_p(1).$$

So we derive

$$S_n(\pi(U), \theta) = \sum_{k=1}^{K} \zeta_k v_k + Z^T \mathbf{u} + \frac{1}{2} \sum_{k=1}^{K} \omega_k g(b^*_{\tau_k})[v_k^2 + \mathbf{u}^T \Sigma_1 \mathbf{u}] + o_p(1).$$

Since $S_n(\pi(U), \theta)$ is a convex function, then following Knight [34] and Koenker [9], we derive

$$\hat{\mathbf{u}}_n \to^d \hat{\mathbf{u}},$$

where $\hat{\mathbf{u}}_n$ and $\hat{\mathbf{u}}$ is the minimizer of $S_n(\pi(U), \mathbf{u}, \mathbf{v})$ and $S(\mathbf{u}, \mathbf{v})$ for $\mathbf{u}$, respectively. By some simple algebra calculation we have

$$\hat{\mathbf{u}} = -[\sum_{k=1}^{K} \omega_k g(b^*_{\tau_k})\Sigma_1]^{-1} Z.$$

Using the fact that $Z \sim N(\mathbf{0}, \Sigma_2(\omega^T \Omega \omega))$, we derive that $\hat{\mathbf{u}}_n \sim N(0, \frac{\omega^T \Omega \omega}{\omega^T \mathbf{g}\mathbf{g}^T \omega} \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1})$.

(2) When the selection probability function is estimated by the nonparametric method, minimizing the object function $L_n(\hat{\pi}(U), \beta, \mathbf{b})$ is equivalent to minimizing $S_n(\hat{\pi}(U), \theta)$. Similar to the proof of (1), we have

$$\begin{aligned}
S^*_n(\hat{\pi}(U), \theta) &= \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(U_i)} \{\frac{1}{\sqrt{n}} \xi_i(\mathbf{u}, v_k)[I(\varepsilon_i < b^*_{\tau_k}) - \tau_k]\} \\
&\quad + \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})] dx \\
&= \sum_{k=1}^{K} \zeta^{(1)}_{nk} v_k + [Z^{(1)}_n]^T \mathbf{u} + \sum_{k=1}^{K} \omega_k B^{(1)}_{nk},
\end{aligned}$$

where

$$\zeta^{(1)}_{nk} = \omega_k \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(U_i)} [I(\varepsilon_i < b^*_{\tau_k}) - \tau_k],$$

$$Z^{(1)}_n = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(U_i)} (\nabla f^*_i)^T \eta_i,$$

$$B^{(1)}_{nk} = \textstyle\sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})] dx.$$

Since

$$\frac{\delta_i}{\hat{\pi}(U_i)} = \frac{\delta_i}{\pi(U_i)} + \frac{\delta_i(\pi(U_i) - \hat{\pi}(U_i))}{\pi(U_i)\hat{\pi}(U_i)},$$

so we have

$$\begin{aligned}
B^{(1)}_{nk} &= B_{nk} + \sum_{i=1}^{n} \frac{\delta_i(\pi(U_i) - \hat{\pi}(U_i))}{\pi(U_i)\hat{\pi}(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})] dx \\
&\hat{=} B_{nk} + \tilde{B}_{nk}.
\end{aligned}$$

Under the Assumption C4–C6, we can show that

$$\sum_{i=1}^{n} \frac{\delta_i}{\pi(U_i)\hat{\pi}(U_i)} \int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u}, v_k)} [I(\varepsilon_i < b^*_{\tau_k} + x) - I(\varepsilon_i < b^*_{\tau_k})] dx = O_p(1).$$

Since $\sup_u |\hat{\pi}(u)-\pi(u)| = o_p(1)$, combine this result with the above one, we can derive that $\tilde{B}_{nk} = o_p(1)$, therefore, $B^{(1)}_{nk} = B_{nk} + o_p(1)$. Similar to the proof of Theroem 3 in Wong et al. [35], we can show that

$$Z_n^{(1)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\frac{\delta_i}{\pi(U_i)}(\nabla f_i^*)^T\eta_i - \frac{\delta_i-\pi(U_i)}{\pi(U_i)}E[(\nabla f_i^*)^T\eta_i|U_i]\} + o_p(1).$$

Noting that ε is independent of X, we can easily derive that $E[(\nabla f_i^*)^T\eta_i|U_i] = E\{E[(\nabla f_i^*)^T\eta_i|X_i]|U_i\} = E\{(\nabla f_i^*)^T E(\eta_i|X_i)|U_i\} = E\{(\nabla f_i^*)^T E(\eta_i)|U_i\} = 0$. So we have $Z_n^{(1)} = Z_n + o_p(1)$. Therefore, we derive that

$$S_n^*(\hat{\pi}(U),\theta) = \sum_{k=1}^{K}\zeta_{nk}^{(1)}v_k + Z_n^T\mathbf{u} + \sum_{k=1}^{K}\omega_k B_{nk} + o_p(1).$$

The following is same to the proof of (1), here we omit. So we complete the proof of (2).

(3) When the selection probability function is estimated by the parametric method, minimizing the object function $L_n(\pi(U,\hat{\gamma}),\beta,\mathbf{b})$ is equivalent to minimizing $S_n(\pi(U,\hat{\gamma}),\theta)$. Using the same argument, we analyse

$$\begin{aligned} S_n^*(\pi(U,\hat{\gamma}),\theta) &= \sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\frac{\delta_i}{\pi(U,\hat{\gamma})}\{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)[I(\varepsilon_i<b^*_{\tau_k})-\tau_k]\} \\ &\quad + \sum_{k=1}^{K}\omega_k\sum_{i=1}^{n}\frac{\delta_i}{\pi(U,\hat{\gamma})}\int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)}[I(\varepsilon_i<b^*_{\tau_k}+x)-I(\varepsilon_i<b^*_{\tau_k})]dx \\ &= \sum_{k=1}^{K}\zeta_{nk}^{(2)}v_k + [Z_n^{(2)}]^T\mathbf{u} + \sum_{k=1}^{K}\omega_k B_{nk}^{(2)}, \end{aligned}$$

where

$$\zeta_{nk}^{(2)} = \omega_k\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(U_i,\hat{\gamma})}[I(\varepsilon_i<b^*_{\tau_k})-\tau_k],$$

$$Z_n^{(2)} = \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\tfrac{\delta_i}{\pi(U_i,\hat{\gamma})}(\nabla f_i^*)^T\eta_i,$$

$$B_{nk}^{(2)} = \textstyle\sum_{i=1}^{n}\tfrac{\delta_i}{\pi(U_i,\hat{\gamma})}\int_0^{\frac{1}{\sqrt{n}}\xi_i(\mathbf{u},v_k)}[I(\varepsilon_i<b^*_{\tau_k}+x)-I(\varepsilon_i<b^*_{\tau_k})]dx.$$

Similarly, the proof will be completed if we can show that

$$S_n^*(\pi(U,\hat{\gamma}),\theta) = \sum_{k=1}^{K}\zeta_{nk}^{(2)}v_k + Z_n^T\mathbf{u} + \sum_{k=1}^{K}\omega_k B_{nk} + o_p(1).$$

Under Assumption C7, we have

$$\frac{\delta_i}{\pi(U_i,\hat{\gamma})} - \frac{\delta_i}{\pi(U_i,\gamma)} = -\frac{\delta_i\frac{\partial\pi(U_i,\gamma)}{\partial\gamma}}{\pi^2(U_i,\gamma)}(\hat{\gamma}-\gamma) + o_p(n^{-1/2}).$$

Noting $\frac{\partial\pi(U_i,\gamma)}{\partial\gamma} = \pi(U_i,\gamma)(1-\pi(U_i,\gamma))\Gamma_i^T$, so we derive

$$\frac{\delta_i}{\pi(U_i,\hat{\gamma})} - \frac{\delta_i}{\pi(U_i,\gamma)} = -\frac{\delta_i(1-\pi(U_i,\gamma))}{\pi(U_i,\gamma)}\Gamma_i^T(\hat{\gamma}-\gamma) + o_p(n^{-1/2}).$$

Similarly, we can verify that $B_{nk}^{(2)} = B_{nk} + o_p(1)$.

$$
\begin{aligned}
Z_n^{(2)} &= Z_n + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(1-\pi(U_i,\gamma))}{\pi(U_i,\gamma)}(\nabla f_i^*)^T\eta_i\Gamma_i(\hat{\gamma}-\gamma)+o_p(1)\\
&= Z_n + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i(1-\pi(U_i,\gamma))}{\pi(U_i,\gamma)}(\nabla f_i^*)^T\eta_i\Gamma_i\sqrt{n}(\hat{\gamma}-\gamma)+o_p(1).
\end{aligned}
$$

Using the fact that ε is independent of X, it is easy to derive

$$
E[\frac{\delta_i(1-\pi(U_i,\gamma))(\nabla f_i^*)^T\eta_i\Gamma_i}{\pi(U_i,\gamma)}] = E\left((1-\pi(U_i,\gamma))(\nabla f_i^*)^T\Gamma_i\right)E(\eta_i) = 0.
$$

So we get

$$
Z_n^{(2)} = Z_n + o_p(1).
$$

With these results in hand, we easily derive

$$
S_n^*(\pi(U,\hat{\gamma}),\theta) = \sum_{k=1}^{K}\zeta_{nk}^{(2)}v_k + Z_n^T\mathbf{u} + \sum_{k=1}^{K}\omega_k B_{nk} + o_p(1).
$$

The following is same to the proof of (1), here we omit. This completes the proof of Theorem 2.1.

The Proof of corollary 2.1 and corollary 2.2 is same to the proof of Proposition 1 and Theorem 2 in Jiang et al. [24], so here we omit.

Proof of Theorem 2.2. (1) When the selection probability function is supposed to be known, from the first part proof of Theorem 2.1, we can derive the Bahadur expression of $\hat{\beta}_{\tau_k}$, because the objective function $\sum_{i=1}^{n}\frac{\delta_i}{\pi(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k})$ with a single quantile is just a special case of Theorem 2.1. Let $\eta_i(\tau_k) = [I(\varepsilon_i < b_{\tau_k}^*) - \tau_k]$ and $Z_{n,\tau_k} = \sqrt{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(U_i)}(\nabla f_i^*)^T\eta_i(\tau_k)$. the Bahadur expression of $\hat{\beta}_{\tau_k}$ is

$$
\sqrt{n}(\hat{\beta}_{\tau_k} - \beta^*) = -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}Z_{n,\tau_k} + o_p(1).
$$

Thus,

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_{OWCQ2} - \beta^*) &= \sqrt{n}(\sum_{k=1}^{K}\omega_k\hat{\beta}_{\tau_k} - \beta^*)\\
&= \sqrt{n}\sum_{k=1}^{K}\omega_k(\hat{\beta}_{\tau_k} - \beta^*)\\
&= \sum_{k=1}^{K}\omega_k\{-\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}Z_{n,\tau_k} + o_p(1)\}\\
&= -E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}\frac{1}{\sqrt{n}}[\sum_{i=1}^{n}\frac{\delta_i}{\pi(U_i)}(\nabla f_i^*)^T\sum_{k=1}^{K}\frac{\omega_k}{g(b_{\tau_k}^*)}\eta_i(\tau_k)] + o_p(1).
\end{aligned}
$$

Using the assumption that ε is independent of X, by some calculations, we can have that

$$
E[\frac{\delta_i}{\pi(U_i)}(\nabla f_i^*)^T\sum_{k=1}^{K}\frac{\omega_k}{g(b_{\tau_k}^*)}\eta_i(\tau_k)] = 0,
$$

and

$$Cov[\frac{\delta_i}{\pi(U_i)}(\nabla f_i^*)^T \sum_{k=1}^{K} \frac{\omega_k}{g(b_{\tau_k}^*)}\eta_i(\tau_k)] = E[\frac{\delta_i}{\pi^2(U_i)}\nabla f_i^*(\nabla f_i^*)^T\{\sum_{k=1}^{K} \frac{\omega_k}{g(b_{\tau_k}^*)}\eta_i(\tau_k)\}^2]$$

$$= E[\frac{\delta_i}{\pi^2(U_i)}\nabla f_1^*(\nabla f_1^*)^T]E[\{\sum_{k=1}^{K} \frac{\omega_k}{g(b_{\tau_k}^*)}\eta_i(\tau_k)\}^2] = \Sigma_2\omega^T H\omega.$$

So we have

$$\sqrt{n}(\hat{\beta}_{OWCQ2} - \beta^*) \xrightarrow{D} N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\omega^T H\omega).$$

(2) When the selection probability function is estimated by the nonparametric method, since the objective function $\sum_{i=1}^n \frac{\delta_i}{\hat{\pi}(U_i)}\rho_{\tau_k}(Y_i - f(X_i,\beta) - b_{\tau_k})$ with a single quantile is just a special case of the second part in the proof of Theorem2.1. Similar to the first part, we can easily derive the Bahadur expression of $\hat{\beta}_{\tau_k,N}$, that is

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{\tau_k,N} - \beta^*) &= -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}Z_{n,\tau_k}^{(1)} + o_p(1) \\
&= -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}\{Z_{n,\tau_k} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i - \pi(U_i)}{\pi(U_i)}E[(\nabla f_i^*)^T\eta_i(\tau_k)|U_i]\} + o_p(1).
\end{aligned}$$

Because ε is independent of X, it is easy to derive that $E[(\nabla f_i^*)^T\eta_i(\tau_k)|U_i] = (\nabla f_i^*)^T E[\eta_i(\tau_k)|U_i] = 0$. Hence, we have

$$\sqrt{n}(\hat{\beta}_{\tau_k,N} - \beta^*) = -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}Z_{n,\tau_k} + o_p(1),$$

which implies the Bahadur expression of $\hat{\beta}_{\tau_k,N}$ is same to that of $\hat{\beta}_{\tau_k}$. So when the selection probability function is estimated by the nonparametric method, we obtain

$$\sqrt{n}(\hat{\beta}_{OWCQ2,N} - \beta^*) \xrightarrow{D} N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\omega^T H\omega).$$

(3) When the selection probability function is estimated by the parametric method, we can obtain the Bahadur expression of $\hat{\beta}_{\tau_k,P}$ from the third part in the proof of Theorem2.1. That is

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{\tau_k,P} - \beta^*) &= -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}Z_{n,\tau_k}^{(2)} + o_p(1) \\
&= -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b_{\tau_k}^*)}\{Z_{n,\tau_k} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i - \pi(U_i,\gamma)}{\pi(U_i,\gamma)}(\nabla f_i^*)^T\eta_i(\tau_k)\Gamma_i(\hat{\gamma} - \gamma)\} + o_p(1).
\end{aligned}$$

Because ε is independent of X, we can easily obtain that

$$\begin{aligned}
E\{\frac{\delta_i - \pi(U_i,\gamma)}{\pi(U_i,\gamma)}(\nabla f_i^*)^T\eta_i(\tau_k)\Gamma_i\} &= E\{\frac{\delta_i - \pi(U_i,\gamma)}{\pi(U_i,\gamma)}(\nabla f_i^*)^T\Gamma_i\}E[\eta_i(\tau_k)] \\
&= E\{(1 - \pi(U_i,\gamma))(\nabla f_i^*)^T\Gamma_i\}E[\eta_i(\tau_k)] \\
&= 0.
\end{aligned}$$

Again we have

$$\sqrt{n}(\hat{\beta}_{\tau_k,P} - \beta^*) = -\frac{E[\nabla f_1^*(\nabla f_1^*)^T]^{-1}}{g(b^*_{\tau_k})} Z_{n,\tau_k} + o_p(1).$$

Therefore,

$$\sqrt{n}(\hat{\beta}_{OWCQ2,P} - \beta^*) \xrightarrow{D} N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\omega^T H\omega).$$

The proof of Theorem 2.2 is completed.