AIMS Mathematics

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

AIMS Mathematics, 9(7): 18117–18133.
DOI:10.3934/math.2024884
Received: 22 March 2024
Revised: 08 May 2024
Accepted: 22 May 2024
Published: 29 May 2024

Research article

Orthogonality based modal empirical likelihood inferences for partially nonlinear models

Jieqiong Lu[1], Peixin Zhao[2] and Xiaoshuang Zhou[3,*]

[1] School of Mathematics and Statistics, Shandong University of Technology, Shandong, Zibo 255000, China
[2] School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China
[3] College of Mathematics and Big Data, Dezhou University, Shandong, Dezhou 253600, China

* **Correspondence:** Email: xiaoshuang1227@126.com.

Abstract: This paper explored the effective empirical likelihood inferences for partially nonlinear models. By combining the modal regression method with orthogonal projection technology, a modal empirical likelihood-based estimation procedure was proposed. The proposed empirical likelihood approach retained Wilk's theorem under mild conditions, and the confidence regions of model coefficients were constructed. Nonparametric and parametric components of the estimators were independent. Simulation results demonstrated that it is more robust and effective than the existing methods.

Keywords: partially nonlinear model; modal regression; empirical likelihood; orthogonality estimation
Mathematics Subject Classification: 62G05, 62G20

1. Introduction

A partially nonlinear model, as a combination of a partially linear model and nonlinear model, takes the following form:

$$Y = g(X, \beta) + \alpha(U) + \varepsilon, \tag{1.1}$$

where $X \in R^p$ and $U \in R$ are covariates, Y is the corresponding response variable, $\beta = (\beta_1, \cdots, \beta_p)^T$ represents a p-dimensional vector comprising unknown regression parameters, and $\alpha(\cdot)$ denotes the unspecified smooth function. $g(\cdot, \cdot)$ is a prespecified function and ε is the model error satisfying $E(\varepsilon|X, U) = 0$. Here, X and β have the same dimension. It is worth noting that this condition is not necessary.

Serving as an extension to both the partially linear model and nonlinear model, model (1.1) has stronger adaptability owing to its ability of to explore more types of nonlinear relationships in data. Recently, many researchers have considered the estimation methods for β and $\alpha(\cdot)$ in model (1.1). For example, Li and Nie [1] developed a simple and effective estimation procedure for β through a nonlinear mixed-effects model. Li and Nie [2] investigated two kinds of estimation methods including profile nonlinear least squares technique and linear approximation approach. Huang and Chen [3] proposed the profile least squares technique, which is based on spline by using graduating functions to approximate the baseline function. Song, Zhao, and Wang [4] introduced a sieve least squares approach tailored to a specific baseline function. Xiao, Tian, and Li [5] carried out the empirical likelihood inferences for model (1.1) and showed its superiority in the establishment of confidence regions. Recently, Que, Huang, and Zhang [6] discovered a partially nonlinear random effects model and provided a new estimation process by combination of B-splines and the generalized least squares approach. Jiang, Tian, and Fei [7] proposed a robust and efficient estimation procedure for model (1.1) with a new MM algorithm.

However, the above-mentioned works mostly paid attention to methods related to least squares or likelihood, which may be susceptible to outliers. Additionally, the estimation efficiency cannot be guaranteed when the model error does not conform to a normal distribution. Due to the prominent merits of quantile regression, Zhou, Zhao, and Gai [8] examined the structure of confidence intervals for partially nonlinear quantile regression models in scenarios where responses are missing at random.

Besides these, modal regression proposed by Yao and Li [9] is an alternative powerful way to ensure the robustness of the estimation procedure through the most likely conditional values. Specifically, using the linear regression model

$$Y_i = X_i^T\beta + \varepsilon_i, i = 1, \cdots n \tag{1.2}$$

as an example to explain the thought of modal regression, assume that $f(y|x)$ be the conditional density function of the response variable Y given the covariate X. Yao and Li [9] defined the modulus of $f(y|x)$ as $\mathrm{Mode}(Y|X) = \arg\max_y(f(y|x))$, and proposed the modal estimator of the unknown parameter vector β by maximizing the subsequent objective function.

$$Q_h(\beta) = \frac{1}{n}\sum_{i=1}^{n} \varphi_h(Y_i - X_i^T\beta), \tag{1.3}$$

where $\varphi_h(\cdot) = h^{-1}\varphi(\cdot/h)$ represents a symmetric kernel function with bandwidth h. Recently, the modal regression method has received increasing interest in applications of several kinds of models. For example, Yao, Lindsay, and Li [10] explored a local modal estimation method for the general nonparametric models. Yao and Li [9] investigated the modal regression method for the linear regression models. Liu et al. [11] studied the statistical inference problem of single index models using modal estimation methods. Zhang, Zhao, and Liu [12] investigated the statistical inferences for varying-coefficient partially linear models based on modal regression. At the same time, Xiao and Liang [13] studied the statistical inferences for varying-coefficient partially nonlinear models based on modal regression. Yang, Lv, and Guo [14] studied the estimation problem of single-index partially linear models using modal regression method. More works are listed but not limited to [15–18]. As explained in these works, the obtained estimators based on modal regression possess not only strong robustness for outliers or heavy-tail error distributions but also high efficiency when the error does not conform to a normal distribution.

As is common knowledge, the empirical likelihood approach, presented by Owen [19], is a very powerful tool in modern statistics, and it has been proven to work well in various nonparametric and semiparametric models under the framework of both the mean regression and the quantile regression. Therefore, the performance of empirical likelihood under modal regression is a worthwhile research topic. Zhao et al. [20] explored a novel empirical likelihood approach based on modal regression and illustrated the robustness and high efficiency. To the best of our knowledge, the empirical likelihood inferences for partially nonlinear models, which are based on modal regression, have not been researched yet.

Considering the aforementioned issues, as presented in this article, we shall pay attention to the empirical likelihood-based estimation for model (1.1) within the structure of modal regression. Employing orthogonal projection technology, we define the auxiliary random vectors for both the parametric vector β and the nonparametric function $\alpha(\cdot)$ for the construction of the empirical log-likelihood ratios, and show that our empirical log-likelihood ratios are asymptotically standard chi-squared distributed. Moreover, simulation studies are performed to test the limited sample performance of the method that we proposed.

The outline of this paper is arranged as below. In Section 2, we mainly pay attention to the parameter vector β, and develop the estimation approach and the asymptotic properties of the estimation that we proposed. In Section 3, we give an account of the estimation result of the nonparametric function $\alpha(\cdot)$. Simulation results are shown in Section 4. A real data example is presented to illustrate our proposed method in Section 5. The conditions and the proof of the primary conclusions are revealed in the Appendix.

2. Empirical likelihood inferences for parametric vector

Let $\{(Y_i, X_i, U_i), i = 1, \ldots, n\}$ be a random sample from the partially nonlinear model below:

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

where $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^p$, and $U_i \in \mathbb{R}$ are the i-th observed values of the variables Y, X, and U, respectively. Due to the presence of both the undetermined parametric vector β and the undetermined nonparametric function $\alpha(\cdot)$, it is complicated to directly derive their estimators because of the different convergence rates. To solve the problem, we intend to parameterize the nonparametric function at first and then use B-spline basis functions to deal with the parametrization of $\alpha(\cdot)$. Particularly, let $B(u) = (B_1(u), \cdots, B_L(u))^T$ be the B-spline basis functions, where $L = K + M$ with the order of M and the number of interior knots K. Thus, $\alpha(u)$ can be approximated as

$$\alpha(u) \approx B(u)^T t, \tag{2.2}$$

where $t = (t_1, \cdots, t_L)^T$ is the vector composed of basis function coefficients. Denote $Z_i = B(U_i) = (B_1(U_i), \cdots, B_L(U_i))^T$, combining (2.1) and (2.2), we obtain

$$Y_i \approx g(X_i,\beta) + Z_i^T t + \varepsilon_i, i = 1, \ldots, n. \tag{2.3}$$

Denote $g(X,\beta) = (g(X_1,\beta), \cdots, g(X_n,\beta))^T$, $Z = (Z_1, \cdots, Z_n)^T$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$ and $Y = (Y_1, \cdots, Y_n)^T$, then the model (2.1) can be rewritten as

$$Y \approx g(X,\beta) + Zt + \varepsilon. \tag{2.4}$$

Here, $Z = Z_{n\times L}$ is presumed to be a matrix of column full rank, and using the QR decomposition technology, Z can be decomposed into

$$Z = Q\begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{2.5}$$

where Q is an $n\times n$ orthogonal matrix, R is an $L\times L$ upper triangular matrix, and 0 is a $(n-L)\times L$ zero matrix. Moreover, we divide Q into $Q = (Q_1, Q_2)$, with the dimensions of Q_1 and Q_2 being $n\times L$ and $n\times(n-L)$, respectively. Since Q is orthogonal, we have $Q_2^T Q_1 = 0$. Therefore, $Q_2^T Z = Q_2^T Q_1 R = 0$ holds. By multiplying Q_2^T on each side of (2.4), we get

$$Q_2^T Y \approx Q_2^T g(X,\beta) + Q_2^T \varepsilon. \tag{2.6}$$

Write $Q_2^T Y = (Y_1^*, \cdots, Y_{n-L}^*)^T$, $Q_2^T g(X,\beta) = (g(X_1^*,\beta), \cdots, g(X_{n-L}^*,\beta))^T$ and $Q_2^T\varepsilon = (\varepsilon_1^*, \cdots, \varepsilon_{n-L}^*)^T$, then model (2.6) can be expressed as

$$Y_i^* = g(X_i^*,\beta) + \varepsilon_i^*, i = 1, \cdots, n-L. \tag{2.7}$$

Model (2.7) can be regarded as a nonlinear model. Inspired by the thought of modal regression and empirical likelihood method, the auxiliary random vector of β can be expressed as

$$\xi_i(\beta) = g'(X_i^*,\beta)\varphi_h'(Y_i^* - g(X_i^*,\beta)), i = 1, \cdots, n-L, \tag{2.8}$$

where $g'(x,\beta) = \frac{\partial g(x,\beta)}{\partial\beta}$ and $\varphi_h'(x)$ is the derivative of the kernel density function $\varphi_h(x)$. Therefore, the empirical log-likelihood ratio of β is as below:

$$R(\beta) = -2\max\left\{\sum_{i=1}^{n-L}\log((n-L)p_i)\middle| p_i \geq 0, \sum_{i=1}^{n-L} p_i = 1, \sum_{i=1}^{n-L} p_i\xi_i(\beta) = 0\right\}.$$

There exists a unique value for $R(\beta)$ where 0 is inside the convex hull formed by $(\xi_1(\beta), \cdots, \xi_{n-L}(\beta))$. Using the Lagrange multiplier method, $R(\beta)$ can be rewritten as

$$R(\beta) = 2\sum_{i=1}^{n-L}\log\{1 + \lambda^T\xi_i(\beta)\}, \tag{2.9}$$

where λ is a Lagrange multiplier that is determined by

$$\frac{1}{n-L}\sum_{i=1}^{n-L}\frac{\xi_i(\beta)}{1+\lambda^T\xi_i(\beta)} = 0. \tag{2.10}$$

The next Theorem 1 shows that the asymptotic distribution of $R(\beta)$ is a standard chi-square distribution, and the detailed proof is included in the Appendix.

Theorem 1. *Assume that Conditions C1–C6 given in the Appendix hold. If β is the true value of the unknown parameter vector, we will have*

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

where $\xrightarrow{D}$ denotes the convergence in distribution.

The confidence region of the parameter vector β can be established based on Theorem 1. More specifically, for a given α $(0 < \alpha < 1)$, c_α is a constant that satisfies $P(\chi_p^2 \leq c_\alpha) = 1 - \alpha$, then $C_\alpha(\beta) = \{\beta | R(\beta) \leq c_\alpha\}$ can be seen as an approximate $(1 - \alpha)$ confidence region of β.

Furthermore, maximizing $\{-R(\beta)\}$ concerning β yields the maximum empirical likelihood estimator, denoted as $\hat{\beta}$, that is

$$\hat{\beta} = \arg\max_{\beta}\{-R(\beta)\}.$$

The next Theorem 2 points out that the limit distribution of $\hat{\beta}$ is asymptotically normal distributed under some regular Conditions $C1$–$C6$ in the Appendix.

Theorem 2. *Assume that Conditions C1–C6 given in the Appendix hold, and the quantity of knots K satisfies $K = O(n^{\frac{1}{2r+1}})$, then as $n \to \infty$, it holds*

$$\sqrt{n-L}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \Gamma^{-1}\Sigma\Gamma^{-1}),$$

where Γ and Σ can be uncovered in $C5$ in the Appendix.

3. Empirical likelihood for nonparametric component

Next, we will study the estimation issues for the unknown $\alpha(u)$ in the model (2.1) based on modal regression. Substitute the unknown β for its maximum empirical likelihood estimator $\hat{\beta}$ into the model (2.1), we get

$$\tilde{Y}_i = Z_i^T t + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $\tilde{Y}_i = Y_i - g(X_i, \hat{\beta})$. Model (3.1) can be regarded as a standard linear model, and $\alpha(u)$ in model (2.1) is linked with t in Eq (2.2). Naturally, we can get the auxiliary random vector for t by

$$\eta_i(t) = Z_i \varphi_h'(\tilde{Y}_i - Z_i^T t), i = 1, \cdots, n. \tag{3.2}$$

Correspondingly, the empirical log-likelihood ratio for t is

$$\ell(t) = -2 \max\left\{ \sum_{i=1}^{n} \log(np_i) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \eta_i(t) = 0 \right\}.$$

There exists a unique value for $\ell(t)$ where 0 is inside the convex hull formed by $(\eta_1(t), \cdots, \eta_n(t))$. Applying the Lagrange multiplier method, $\ell(t)$ can be rewritten as

$$\ell(t) = 2 \sum_{i=1}^{n} \log\{1 + l^T \eta_i(t)\}, \tag{3.3}$$

where l is the Lagrange multiplier, it can be ascertained by the equation

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\eta_i(t)}{1 + l^T \eta_i(t)} = 0. \tag{3.4}$$

Denote $\hat{t}$ as the solution of maximizing -$\ell(t)$, that is

$$\hat{t} = \arg\max_{t}\{-\ell(t)\},$$

then the estimator of $\alpha(u)$ can be denoted as $\hat{\alpha}(u) = B(u)^T \hat{t}$.
The following Theorem 3 provides the consistency of the estimator $\hat{\alpha}(u)$ and gives its convergence rate.

Theorem 3. *Assume that Conditions C1–C6 in the Appendix hold, and the quantity of interior knots K satisfies $K = O_p(n^{\frac{1}{2r+1}})$, then as $n \to \infty$, it holds*

$$\|\hat{\alpha}(u) - \alpha(u)\| = O_p(n^{\frac{-r}{2r+1}}),$$

where $\|\cdot\|$ denotes the L_2 norm.

Remark. In addition, in practical applications, the quantity of interior knots K and bandwidth h should be chosen properly. We present a chosen method that is suggested by [7]. Define $\hat{F}(h,K) = \frac{1}{n}\sum_{i=1}^{n}\varphi_h''(\hat{\varepsilon}_i)$ and $\hat{G}(h,K) = \frac{1}{n}\sum_{i=1}^{n}\varphi_h'(\hat{\varepsilon}_i)^2$, where $\hat{\varepsilon}_i = Y_i - g(X_i,\hat{\beta}) - \hat{\alpha}(U_i)$. Then, the estimators of K and h can be illustrated by

$$(\hat{h},\hat{K}) = \arg\max_{h,K}\left\{\frac{\hat{F}(h,K)^2}{\hat{G}(h,K)}\right\}. \tag{3.5}$$

Although the parameters given based on this method may not be optimal in theory, the data simulation results in Section 4 point out that our approach is feasible.

4. Simulation studies

To illustrate the finite sample performance of the method that we proposed in this paper, many simulation experiments are conducted in this section. For these simulation experiments, we make use of the kernel function $K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1)$. We need the undersmoothing bandwidth in our estimation process, and the bandwidth h is selected by the "leave-one-out" cross-validation method. Using similar arguments to [21], we can confirm that bandwidth satisfies $h = O(n^{-1/5})$. The data are produced from the following model:

$$Y_i = \exp(X_i^T\beta) + \alpha(U_i) + \varepsilon_i, i = 1, \cdots, n, \tag{4.1}$$

where the parameter $\beta = (\beta_1,\beta_2)^T$ is taken as $\beta = (1,2)^T$, and the nonparametric function $\alpha(u) = \sin(2\pi u)$. The covariates $X_i = (X_{1i}, X_{2i})^T$ and U_i are taken as $X_{1i} \sim N(1,1)$, $X_{2i} \sim N(1,1)$ and $U_i \sim \mathrm{U}[0,1]$, respectively. The response Y_i is produced according to the model, and the model error ε_i is taken from $\varepsilon_i = \varepsilon_i^* + \delta_i e_i$, where $\varepsilon_i^* \sim N(0, 0.5^2)$, $e_i \sim N(0,1)$, and δ_i follows the two-point distribution with $P(\delta_i = 1) = 0.1$ and $P(\delta_i = 0) = 0.9$. Obviously, $\delta_i = 1$ implies the corresponding Y_i is the outlier, and this can make sure that the generated data contains approximately 10% outliers. In addition, as shown in the simulations below, we set the sample size that is $n = 200$, 400, and 600, and for each situation, the experiment is carried 1000 times individually.

To solve the estimation question of nonlinear and parametric components of the partially nonlinear model, Li and Nie [2] proposed a profile nonlinear least squares estimation (PLE) technique. There are two stages for the estimation procedure: at the initial stage, the unknown function is estimated using the local polynomial method provided that the parameters are known, and in the subsequent stage, the parameters that are unknown are estimated using the profile nonlinear least squares method. Next, for this simulation process, we proposed a modal empirical likelihood (MEL) method that will be contrasted with the profile nonlinear least squares estimation (PLE) method proposed by Li and Nie [2].

Based on 1000 simulation runs, the box plots present the 1000 absolute deviation values for parameters β_1 and β_2 that we can see in Figures 1 and 2, defined by $|\hat{\beta}_k - \beta_k|$, $k = 1, 2$. It is obvious that

Figures 1(a) and 2(a) show the simulation results that are based on the MEL method, and Figures 1(b) and 2(b) show the simulation results that are based on the PLE method. From Figures 1 and 2, it is evident that both absolute deviation values, gained by the MEL and PLE methods, decrease with the increase of the sample size. Also, it can be observed that the MEL method outperforms the PLE method, because for any given sample size n, the MEL method can give smaller absolute deviation uniformly. These results indicate that the PLE method is sensitive to outliers, and the proposed MEL method in this paper can trim off the influence of outliers. Therefore, it is more robust.

Next, the performance of confidence regions for parametric components $\beta = (\beta_1, \beta_2)^T$ is evaluated. Based on 1000 simulation runs, as shown in Figure 3, the 95% confidence regions of β and the corresponding coverage probabilities are shown. The dotted curve represents the results based on the MEL method, and the dashed curve the results based on the PLE method. As shown in Figure 3, the confidence regions gained by the MEL method based are better than those gained by the PLE method. This is because the MEL method can provide smaller confidence regions than the PLE method. This is mainly affected by the disadvantages of the PLE method. The confidence region obtained by the PLE method needs to estimate the asymptotic variance of $\hat{\beta}$, so it will make a difference to the accuracy of the resulting confidence region. On the contrary, the MEL-based confidence region has nothing to do with any asymptotic variance estimation and holds a relatively higher accuracy.

Lastly, we evaluate the performance of the confidence intervals for nonparametric function $\alpha(u)$ gained by MEL and PLE methods individually. When the confidence intervals are constructed for PLE estimator by using normal approximation, the asymptotic variance of $\hat{\alpha}(u)$ also needs to be estimated. However, due to the complexity of estimating the asymptotic variance of $\hat{\beta}$, it is difficult to obtain a accurate estimation. Next, we estimate the asymptotic variance in this simulation using the bootstrap method. Under the situation of $n = 400$, with the 1000 simulation runs, the average length of 95% pointwise confidence intervals and the corresponding coverage probabilities for $\alpha(u)$ are computed. We can see the results that are presented in Figures 4 and 5, individually. The dotted curves give the results based on the MEL method; at the same time, the dashed curves give the results based on the PLE method, and the real curve of $\alpha(u)$ is presented by the solid curve.

As shown in Figures 4 and 5, it is noticeable that the MEL method also performs better than the PLE method for the estimation for the nonparametric function $\alpha(u)$; this is because the associated pointwise confidence intervals based on the MEL method have shorter average length, and the corresponding coverage probabilities are a little higher than that gained based on the PLE method. Owing to the estimation of asymptotic covariance, it is usually used in the normal approximation, and will make an influence in the accuracy of the PLE method-based confidence intervals. In contrast, the MEL method does not have to offer a plug-in estimator of the limiting variance. Also, empirical likelihood-based confidence intervals using modal methods can more accurately capture the true shape of the confidence intervals. Compared to the PLE method (the two-stage estimation method), it is remarkable that our proposed MEL method can significantly reduce the computational complexity by using the orthogonal projection technology.

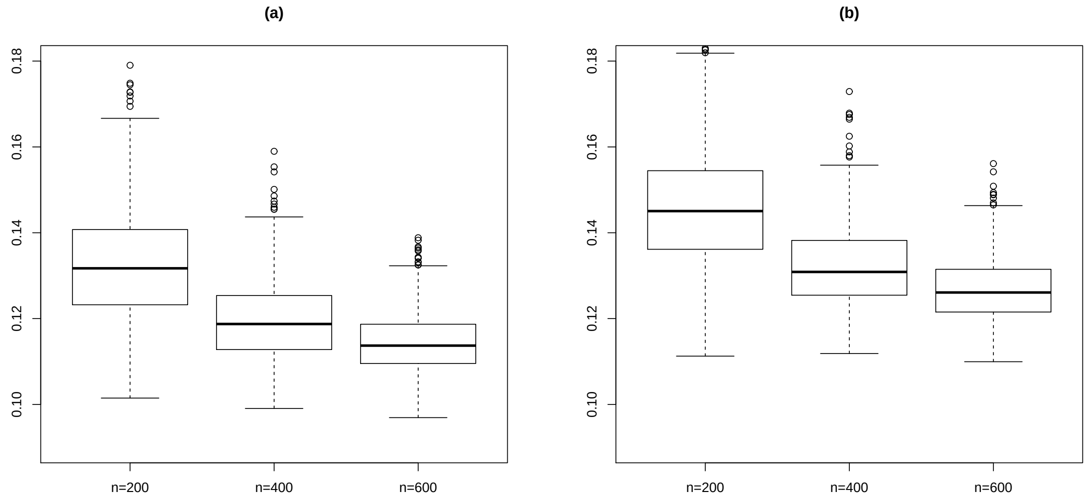

Figure 1. The boxplots of the 1000 absolute deviation values for the estimator of β_1 based on the MEL method (a) and the PLE method (b).

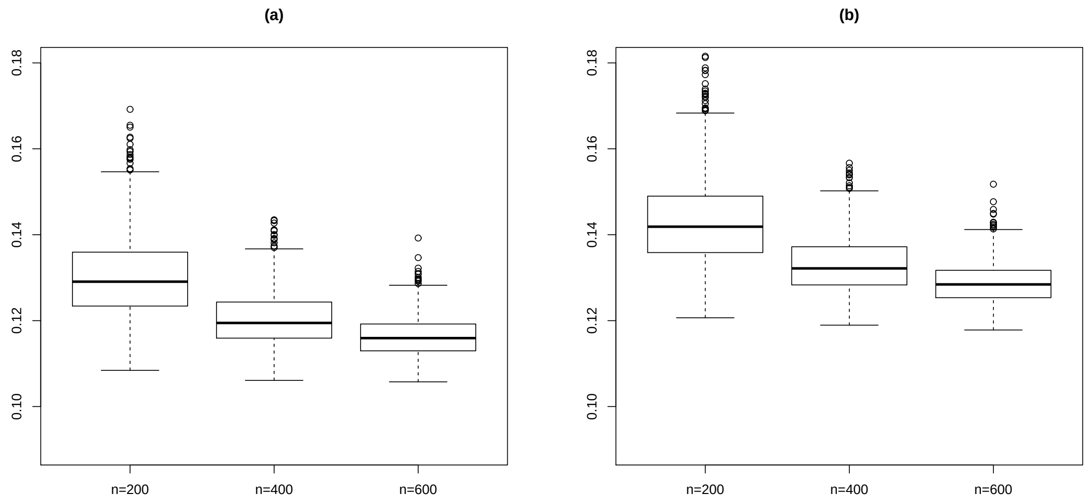

Figure 2. The boxplots of the 1000 absolute deviation values for the estimator of β_2 based on the MEL method (a) and the PLE method (b).

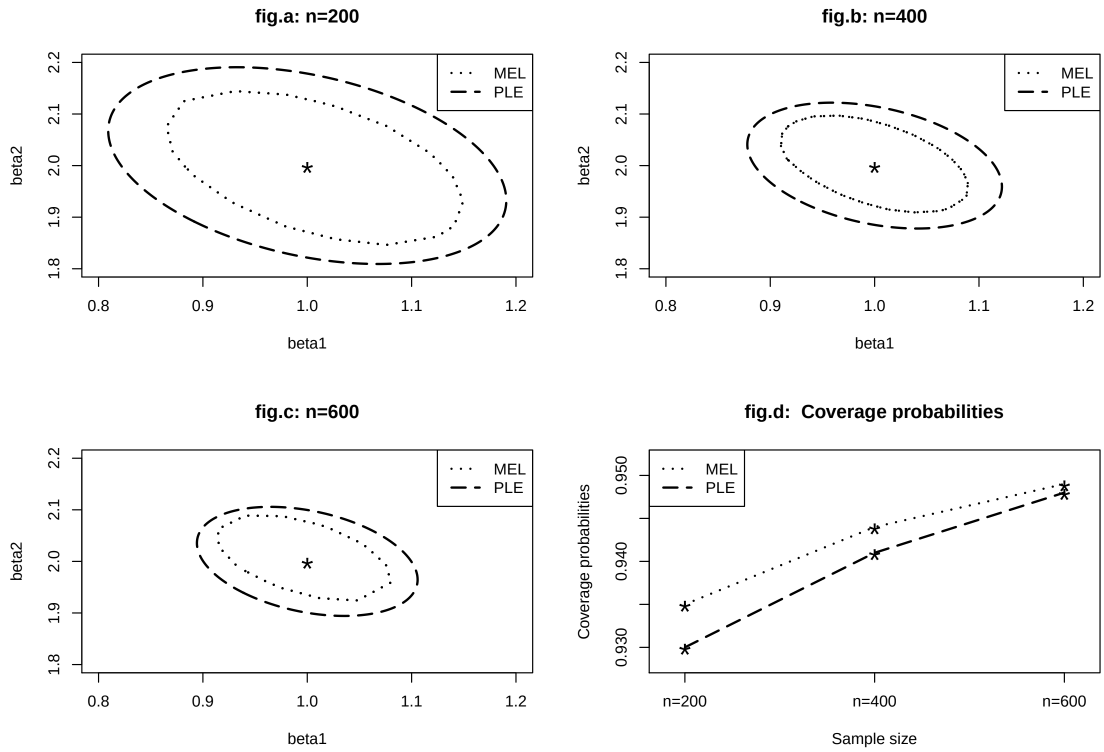

Figure 3. The 95% confidence regions for (β_1, β_2) (a-c) and the corresponding coverage probabilities(d) based on MEL and PLE methods.

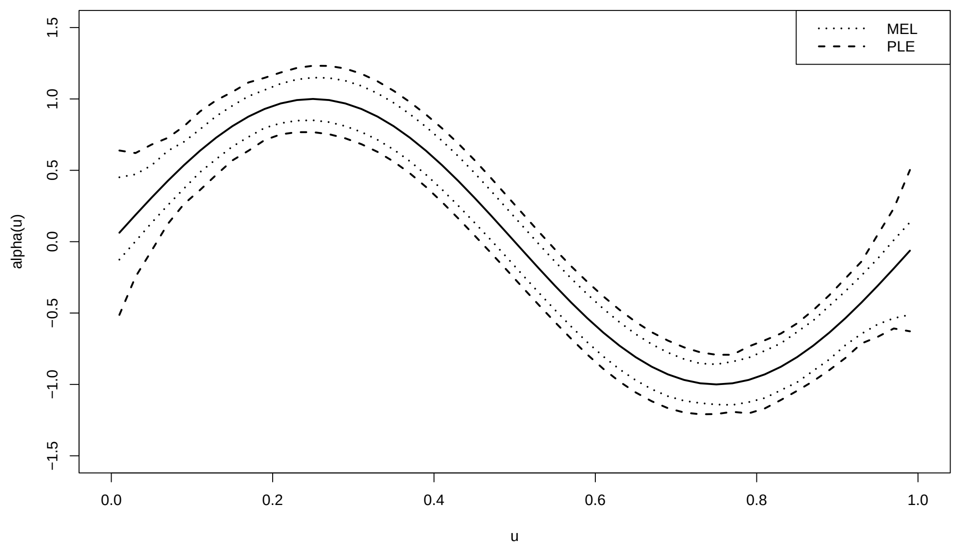

Figure 4. The 95% pointwise confidence intervals for $\alpha(u)$ based on MEL and PLE methods.

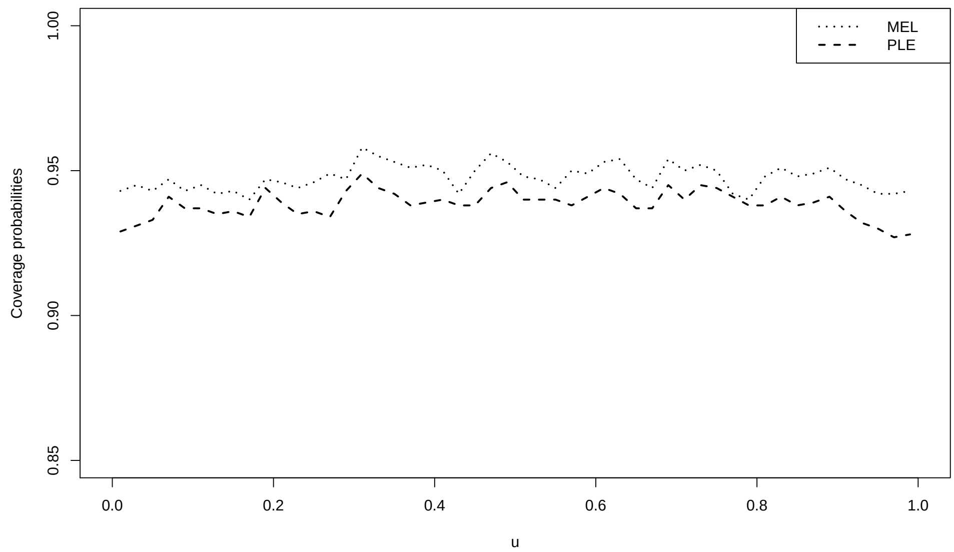

Figure 5. The coverage probabilities of 95% pointwise confidence intervals for $\alpha(u)$ based on MEL and PLE methods.

5. A real data example

In this section, we illustrate the application of the proposed modal empirical likelihood (MEL) method by analyzing the Boston housing price data, consisting of 506 observations from Boston Standard Metropolitan Statistical Area in 1970. This data have also been studied in Li and Mei [22], Zhou, Zhao and Wang [23], and Wang, Zhao and Du [24]. To illustrate the proposed method, we analyze this data by using the following partially nonlinear model

$$Y = \exp(X_1\beta_1 + X_2\beta_2) + g(U) + \varepsilon, \tag{5.1}$$

where the response variable Y denotes the median value of owner-occupied homes in 1000, X_1 denotes the index of accessibility to radial highways, X_2 denotes the pupil-teacher ratio by town school district, and U denotes the square root of the proportion of population that is in the lower status.

In the following estimation procedure, the kernel function is taken by $K(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \leq 1)$, and the bandwidth is taken by (3.5). In addition, for comparison, we also present the estimation results obtained by the profile nonlinear least squares estimation (PLE) method. The estimator of $(\beta_1, \beta_2)^T$ based on our MEL method is $(0.3776, 0.2216)^T$, and the estimator based on the PLE method is $(0.3694, 0.2141)^T$, whose signs are consistent with that obtained by our MEL method, but the values are slightly lower. This potentially indicates that, if we ignore outliers, we would underestimate the impact of X_1 and X_2 on the response Y.

The 95% confidence region of $(\beta_1, \beta_2)^T$, obtained by the PLE and MEL methods, are shown in Figure 6, where the dotted curve represents the results based on the MEL method, and the dashed curve represents the results based on the PLE method. Figure 6 shows that the MEL method can provide smaller confidence regions than the PLE method, which also validates the conclusions obtained from the simulation.

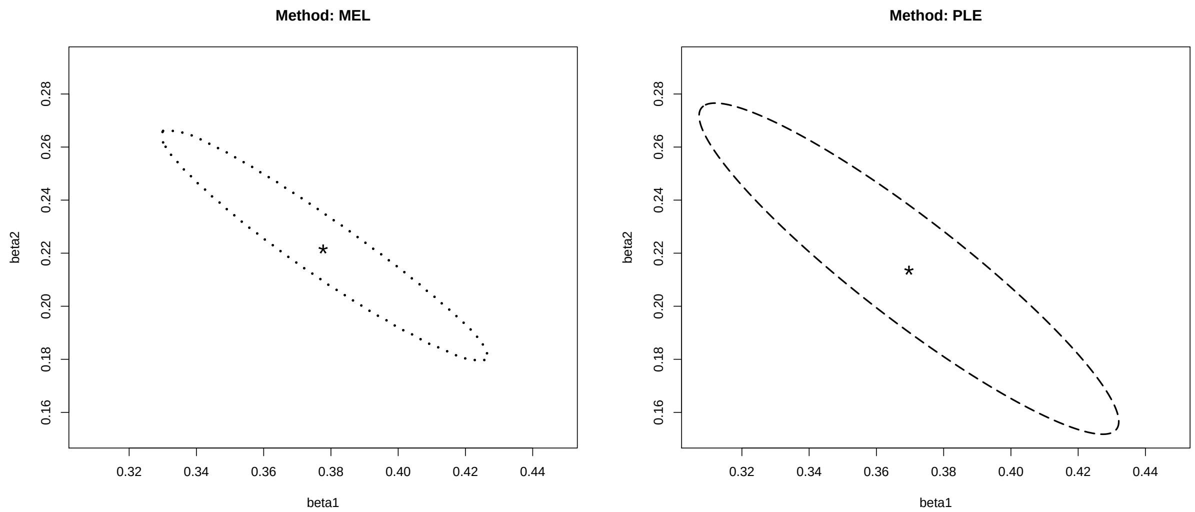

Figure 6. The estimator and the 95% confidence region of (β_1, β_2) based on the MEL and PLE methods.

In addition, the pointwise confidence intervals for the nonparametric component $g(u)$ with nominal level 95% are shown in Figure 7. It can be seen from Figure 7 that the estimated coefficient functions for nonparametric component based on MEL and PLE are very similar. However, the interval width given by the MEL method is slightly shorter than that given by the PLE method, which further confirms that the proposed MEL method is preferable in real data analysis.

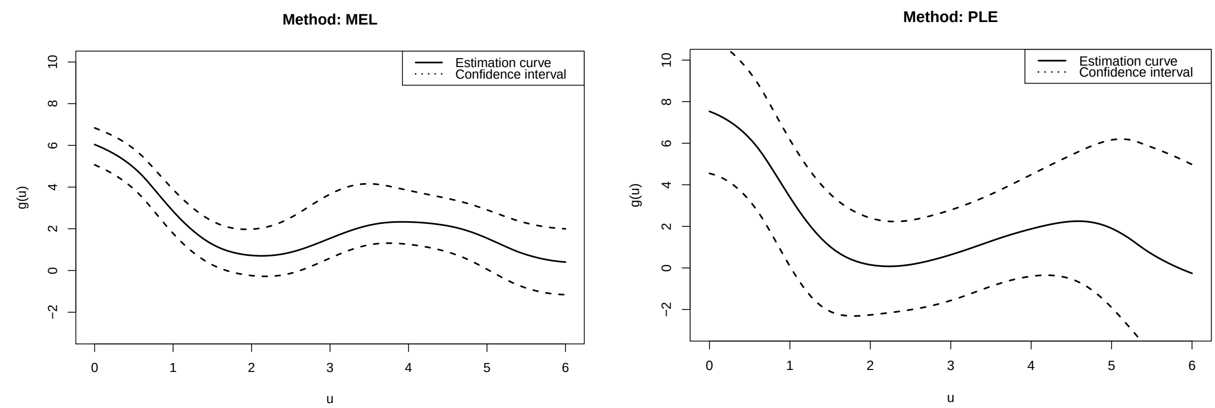

Figure 7. The estimation curve and the 95% confidence interval of $g(u)$ based on the MEL and PLE methods.

6. Discussion and conclusions

In this article, we focus on the empirical likelihood inferences for the partially nonlinear models based on the modal regression approach by means of B-spline basis functions to deal with the parametrization of the nonparametric function, and the relevant asymptotic properties of the resulting estimators under some mild conditions are obtained. Meanwhile, the corresponding confidence regions of parametric and nonparametric components are constructed. Numerical studies show that our modal empirical likelihood method has better advantages than the conventional profile nonlinear least squares method. The main reason is that when there are some large outliers in the data, the modal regression approach will put more weights on the most likely data around the true value, which leads to a robust and efficient estimator.

In addition, our proposed method can be attempted to be extended to other semiparametric nonlinear models, such as varying-coefficient partially nonlinear model, additive partially nonlinear model, and others. Furthermore, robust estimation for the partially nonlinear models and other semiparametric nonlinear models under the framework of longitudinal data is an interesting topic, and we will consider this important topic in our future work.

Author contributions

Jieqiong Lu: Writing-original draft, Conceptualization, Formal analysis, Methodology; Peixin Zhao: Funding acquisition, Software, Supervision, Writing-review and editing; Xiaoshuang Zhou: Funding acquisition, Validation and data analysis, Methodology, Supervision, Writing-review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The research is supported by the Natural Science Foundation of Shandong Province (Grant Nos. ZR2020MA021 and ZR2022MA065).

Conflict of interest

The author declares no conflicts of interest in this paper.

References

1. R. Li, L. Nie, A new estimation procedure for partially nonlinear model via mixed effects approach, *Can. J. Stat.*, **35** (2007), 399–411. https://doi.org/10.1002/cjs.5550350305
2. R. Li, L. Nie, Efficient statistical inference procedures for partially nonlinear models and their applications, *Biometrics*, **64** (2008), 904–911. https://doi.org/10.1111/j.1541-0420.2007.00937.x
3. T. Huang, H. Chen, Estimating the parametric component of nonlinear partial spline model, *J. Multivariate Anal.*, **99** (2008), 1665–1680. https://doi.org/10.1016/j.jmva.2008.01.007
4. L. Song, Y. Zhao, X. Wang, Sieve least squares estimation for partially nonlinear models, *Stat. Probabil. Lett.*, **80** (2010), 1271–1283. https://doi.org/10.1016/j.spl.2010.04.006

5. Y. Xiao, Z. Tian, F. Li, Empirical likelihood-based inference for parameter and nonparametric function in partially nonlinear models, *J. Korean Stat. Soc.*, **43** (2014), 367–379. http://dx.doi.org/10.1016/j.jkss.2013.11.002
6. Y. Que, Z. Huang, R. Zhang, Statistical estimation in partially nonlinear models with random effects, *Statistical Theory and Related Fields*, **1** (2017), 227–233. https://doi.org/10.1080/24754269.2017.1396425
7. Y. Jiang, G. Tian, Y. Fei, A robust and efficient estimation method for partially nonlinear models via a new MM algorithm, *Stat. Papers*, **60** (2019), 2063–2085. https://doi.org/10.1007/s00362-017-0909-5
8. X. Zhou, P. Zhao, Y. Gai, Imputation-based empirical likelihood inferences for partially nonlinear quantile regression models with missing responses, *AStA Adv. Stat. Anal.*, **106** (2022), 705–722. https://doi.org/10.1007/s10182-022-00441-z
9. W. Yao, L. Li, A new regression model: modal linear regression, *Scand. J. Stat.*, **41** (2014), 656–671. https://doi.org/10.1111/sjos.12054
10. W. Yao, B. Lindsay, R. Li, Local modal regression, *J. Nonparametr. Stat.*, **24** (2012), 647–663. https://doi.org/10.1080/10485252.2012.678848
11. J. Liu, R. Zhang, W. Zhao, Y. Lv, A robust and efficient estimation method for single index models, *J. Multivariate Anal.*, **122** (2013), 226–238. https://doi.org/10.1016/j.jmva.2013.08.007
12. R. Zhang, W. Zhao, J. Liu, Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression, *J. Nonparametr. Stat.*, **25** (2013), 523–544. https://doi.org/10.1080/10485252.2013.772179
13. Y. Xiao, L. Liang, Robust estimation and variable selection for varying-coefficient partially nonlinear models based on modal regression, *J. Korean Stat. Soc.*, **51** (2022), 692–715. https://doi.org/10.1007/s42952-021-00158-w
14. H. Yang, J. Lv, C. Guo, Robust estimation and variable selection for varying-coefficient single-index models based on modal regression, *Commun. Stat.-Theor. M.*, **45** (2016), 4048–4067. https://doi.org/10.1080/03610926.2014.915043
15. Y. Xia, Y. Qu, N. Sun, Variable selection for semiparametric varying coefficient partially linear model based on modal regression with missing data, *Commun. Stat.-Theor. M.*, **48** (2019), 5121–5137. https://doi.org/10.1080/03610926.2018.1508712
16. X. Sun, K. Wang, S. Li, L. Lin, Modal regression based robust empirical likelihood and variable selection for partial linear models with longitudinal data (Chinese), *Scientia Sinica (Mathematica)*, **52** (2022), 447–466. https://doi.org/10.1360/SSM-2019-0230
17. Y. Zhou, R. Mei, Y. Zhao, Z. Hu, M. Zhao, Orthogonality-based bias-corrected empirical likelihood inference for partial linear varying coefficient EV models with longitudinal data, *J. Comput. Appl. Math.*, **443** (2024), 115751. https://doi.org/10.1016/j.cam.2023.115751
18. W. Zhao, R. Zhang, J. Liu, Y. Lv, Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression, *Ann. Inst. Stat. Math.*, **66** (2014), 165–191. https://doi.org/10.1007/s10463-013-0410-4

19. A. Owen, Empirical likelihood ratio confidence regions, *Ann. Statist.*, **18** (1990), 90–120. https://doi.org/10.1214/aos/1176347494

20. W. Zhao, R. Zhang, Y. Liu, J. Liu, Empirical likelihood based modal regression, *Stat. Papers*, **56** (2015), 411–530. https://doi.org/10.1007/s00362-014-0588-4

21. L. Xue, L. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, *Biometrika*, **94** (2007), 921–937. https://doi.org/10.1093/biomet/asm066

22. T. Li, C. Mei, Estimation and inference for varying coefficient partially nonlinear models, *J. Stat. Plan. Infer.*, **143** (2013), 2023–2037. https://doi.org/10.1016/j.jspi.2013.05.011

23. X. Zhou, P. Zhao, X. Wang, Empirical likelihood inferences for varying coefficient partially nonlinear models, *J. Appl. Stat.*, **44** (2017), 474–492. https://doi.org/10.1080/02664763.2016.1177496

24. X. Wang, P. Zhao, H. Du, Statistical inferences for varying coefficient partially non linear model with missing covariates, *Commun. Stat.-Theor. M.*, **50** (2021), 2599–2618. https://doi.org/10.1080/03610926.2019.1674870

Appendix. Proofs of main results

We start by outlining the necessary assumption conditions for deriving the main results. These conditions are relatively lenient and can be readily met.

C1. The covariate U has a bounded support, and its density function $f(\cdot)$ is Lipschitz continuous and bounded away from 0 on its support.

C2. The nonparametric function $\alpha(\cdot)$ is r-th continuously differentiable on the interval $(0, 1)$ with $r \geq 2$.

C3. For any x, both $g(x,\beta)$ and the r-th derivative of $g(x,\beta)$ concerning β are continuous functions of β.

C4. Denote $d_1, \cdots, d_K$ be the interior knots in $[0, 1]$. Let $d_0 = 1$, $d_{K+1} = 1$ and $\Delta_i = d_i - d_{i-1}$. Then, there exists a constant C such that

$$\frac{\max \Delta_i}{\min \Delta_i} \leq C, \quad \max \mid \Delta_{i+1} - \Delta_i \mid = o(K^{-1}).$$

C5. For any given bandwidth h, the kernel function $\varphi_h(\cdot)$ satisfies $E\varphi'_h(\varepsilon) = 0$, $E\varphi''_h(\varepsilon) = 0$ and $E\varphi'_h(\varepsilon)^2 < \infty$.

C6. Write $\Sigma = E\varphi'_h(\varepsilon^*_i)g'(X^*_i,\beta)g'(X^*_i,\beta)^T$ and $\Gamma = E\varphi''_h(\varepsilon^*_i)g'(X^*_i,\beta)g'(X^*_i,\beta)^T$, both Σ and Γ are positive matrices.

Lemma 1. *Suppose that $\xi_i, i = 1, \cdots, n$ be the independent random variable series, $E(\xi_i) = 0$, $E(\xi_i^2) < \infty$, then we have*

$$\max_{1\leq k\leq n} \mid \sum_{i=1}^{k} \xi_i \mid = O_p(\sqrt{n} \log n).$$

The proof can be referred to [9].

Lemma 2. *Suppose that Conditions C1–C6 hold, we have*

$$\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}\xi_i(\beta)\xrightarrow{\mathcal{L}}N(0,\Sigma),$$

where Σ *has been defined in Condition C6.*

Proof. Let $R(U_i)=\alpha(U_i)-B(U_i)^Tt$ and $R(U)=(R(U_1),\cdots,R(U_{n-L}))^T$, since $Q_2^TZ=0$, we can get

$$Q_2^TY=Q_2^Tg(X,\beta)+Q_2^TZt_0+Q_2^TR(U)+Q_2^T\varepsilon=Q_2^Tg(X,\beta)+Q_2^TR(U)+Q_2^T\varepsilon. \tag{A.1}$$

Note $Q_2^TR(U)=(r_1^*(U),\cdots,r_{n-L}^*(U))^T$, then invoking (A.1) and the definition of $\xi_i(\beta)$, by Taylor expansion with simple calculations, we can have

$$\begin{aligned}\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}\xi_i(\beta) &= \frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(r_i^*(U)+\varepsilon_i^*)\\ &= \frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(\varepsilon_i^*)+\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h''(\varepsilon_i^*)r_i^*(U)\\ &+ O_p((n-L)^{1/2})\|R(U)\|^2\\ &\equiv J_{n1}+J_{n2}+O_p((n-L)^{1/2}\|R(U)\|^2).\end{aligned} \tag{A.2}$$

It is easy to obtain $E(J_{n1})=0$ and $Var(J_{n1})=\Sigma$, by means of the central limit theorem, we get

$$J_{n1}\xrightarrow{\mathcal{L}}N(0,\Sigma). \tag{A.3}$$

Moreover, by the Conditions $C1$, $C3$, and Proposition 6.21 of [8], we obtain $\|R(U)\|=O(K^{-r})$, then combine Lemma 1 and some simple calculations, we get $J_{n2}=O_p((n-L)^{-\frac{1}{2}})(n-L)^{\frac{1}{2}}\log(n-L)K^{-r}=o_p(1)$ and $O_p((n-L)^{\frac{1}{2}}\|R(U)\|^2)=o_p(1)$. Then, combine (A.2), (A.3), and the Slutsky's theorem, it follows that $\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}\xi_i(\beta)\xrightarrow{\mathcal{L}}N(0,\Sigma)$.

The proof of Lemma 2 is finished. □

Proof of Theorem 1. With the similar arguments of [16], invoking the definition of $\xi_i(\beta)$, we can easily derive

$$\max_i\|\xi_i(\beta)\|=o_p((n-L)^{1/2}), \tag{A.4}$$

and

$$\|\lambda\|=O_p((n-L)^{-1/2}). \tag{A.5}$$

Furthermore, combining (A.3) with (A.4), the following conclusion can be obtained,

$$R(\beta)=\left\{\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}\xi_i(\beta)\right\}^T\hat{\Lambda}^{-1}\left\{\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}\xi_i(\beta)\right\}+o_p(1), \tag{A.6}$$

where $\hat{\Lambda}=(n-L)^{-1}\sum_{i=1}^{n-L}\xi_i(\beta)\xi_i^T(\beta)$, by the law of large numbers, $\hat{\Lambda}\xrightarrow{\mathcal{L}}\Sigma$ holds. By (A.5) and Lemma 2, the proof of Theorem 1 is completed. □

Proof of Theorem 2. It is easy to understand that the maximum empirical likelihood estimator $\hat{\beta}$ is the solution of the estimation equation $\sum_{i=1}^{n-L}\xi_i(\beta)=0$. Hence, we get

$$\sum_{i=1}^{n-L}\xi_i(\hat{\beta})=\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(Y_i^*-g(X_i^*,\hat{\beta}))=0. \tag{A.7}$$

By means of Taylor expansion, we conclude that

$$\begin{aligned}\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(Y_i^*-g(X_i^*,\hat{\beta}))^T &= \sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(\varepsilon_i^*)+\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h''(\varepsilon_i^*)g'(X_i^*,\beta)^T(\beta-\hat{\beta})\\ &+ \sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h''(\varepsilon_i^*)r_i^*(U)+O_p(n\cdot n^{-\frac{2r}{2r+1}}).\end{aligned} \tag{A.8}$$

Combine (A.7) and (A.8) , using some simple calculations, we derive

$$\frac{1}{n-L}\sum_{i=1}^{n-L}g'(X_i^*,\beta)g'(X_i^*,\beta)^T\varphi_h''(\varepsilon_i^*)\sqrt{n-L}(\beta-\hat{\beta})=\frac{1}{\sqrt{n-L}}\sum_{i=1}^{n-L}g'(X_i^*,\beta)\varphi_h'(\varepsilon_i^*)+o_p(1). \tag{A.9}$$

Applying the law of large numbers,

$$\frac{1}{n-L}\sum_{i=1}^{n-L}g'(X_i^*,\beta)g'(X_i^*,\beta)^T\varphi_h''(\varepsilon_i^*)\xrightarrow{\mathcal{L}}\Gamma. \tag{A.10}$$

Then, by the proof of Lemma 2, (A.9), and (A.10), applying the Slutsky Lemma , we can conclude that $\sqrt{n-L}(\hat{\beta}-\beta)\xrightarrow{\mathcal{L}}N(0,\Gamma^{-1}\Sigma\Gamma^{-1})$, which completes the proof of Theorem 2. □

Proof of Theorem 3. By a similar proof of Theorem 2, the maximum empirical likelihood estimator $\hat{t}$ of t is the solution of the equation

$$\sum_{i=1}^{n}\eta_i(t)=\sum_{i=1}^{n}Z_i\varphi_h'(\tilde{Y}_i-Z_i^Tt)=0.$$

Then, it is easy to get that $\hat{t}$ is also the solution of minimizing the following objective function

$$Q_h(t)=\sum_{i=1}^{n}\varphi_h(\tilde{Y}_i-Z_i^Tt). \tag{A.11}$$

Denote $\tau=n^{-\frac{r}{2r+1}}$ and $t=t_0+\tau\xi$, where t_0 is the true value of the parameter vector t. We first demonstrate that for any given $\varepsilon>0$, there exists a large enough constant C such that

$$P\left\{\inf_{\|\xi\|=C}Q_h(t)>Q_h(t_0)\right\}\geq 1-\varepsilon. \tag{A.12}$$

Using Taylor expansion, some calculations yield

$$Q_h(t)=Q_h(t_0+\tau\xi)=Q_h(t_0)+\tau\xi^TQ_h'(t_0)+\frac{1}{2}\tau^2\xi^TQ_h''(t_0)\xi+\frac{1}{2}\tau^3\left(\frac{\partial\{\xi^TQ_h''(\tilde{t})\xi\}}{\partial t}\right)^T\xi, \tag{A.13}$$

where $\tilde{t}$ is between t_0 and t. Moreover, $\hat{\beta} - \beta = O_p(n^{-1/2})$. Combining the proof process of Theorem 2, we obtain

$$\tau\xi^T Q_h'(t_0) = -2\tau\xi^T \left\{ \sum_{i=1}^{n} Z_i \varphi_h'(g(X_i,\beta) - g(X_i,\hat{\beta})) + R(U_i) + \varepsilon_i \right\} = O_p(n^{1/2})\|\xi\|. \tag{A.14}$$

$$\begin{aligned} \frac{1}{2}\tau^2\xi^T Q_h''(t_0)\xi &= \frac{1}{2}\tau^2\xi^T \left\{ 2\sum_{i=1}^{n} Z_i Z_i^T \varphi_h''(g(X_i,\beta) - g(X_i,\hat{\beta})) + R(U_i) + \varepsilon_i \right\}\xi \\ &= O_p(n\tau^2)\xi^T E\{\varphi_h''(\varepsilon_i) Z_i Z_i^T\}\xi + \|\xi\|^2 o_p(n\tau^2). \end{aligned} \tag{A.15}$$

Note $\Omega(t) = K^{-1}(Q_h(t) - Q_h(t_0))$, by the (A.14) and (A.15), we have

$$\begin{aligned} \Omega(t) &= O_p(n\tau^2 K^{-1})\xi^T E\{\varphi_h''(\varepsilon_i) Z_i Z_i^T\}\xi + \|\xi\| O_p(n^{\frac{1}{2}}\tau K^{-1}) + \|\xi\|^2 o_p(n\tau^2 K^{-1}) \\ &\equiv J_1 + J_2 + J_3. \end{aligned} \tag{A.16}$$

Since $K = O_p(n^{\frac{1}{2r+1}})$, we get $O_p(n\tau^2 K^{-1}) = O_p(1)$, $O_p(n^{\frac{1}{2}}\tau K^{-1}) = o_p(1)$ and $O_p(n\tau^2 K^{-1}) = o_p(1)$. Therefore, for a sufficiently large constant C, J_1 dominates J_2 uniformly, and J_1 dominates J_3 uniformly too. With the help of $J_1 > 0$, (A.12) holds naturally. So there exists a local minimizer $\hat{t}$ satisfying

$$\|\hat{t} - t_0\| = O_p(\tau) = O_P(n^{-\frac{r}{2r+1}}). \tag{A.17}$$

Then we have

$$\|\hat{\alpha}(u) - \alpha(u)\|^2 = O_p(\|\hat{t} - t_0\|^2) + O_p(\|R(u)\|^2). \tag{A.18}$$

$O_p(\|R(u)\|^2) = O_P(n^{-\frac{r}{2r+1}})$ can be proved from $\|R(u)\| = O(K^{-r})$ and $K = O_p(n^{\frac{1}{2r+1}})$. Then, by the above two formulas, we get $\|\hat{\alpha}(u) - \alpha(u)\|^2 = O_P(n^{-\frac{2r}{2r+1}})$. The proof of Theorem 3 is completed. □

AIMS Press