Research Article

Empirical Likelihood for Partially Linear Single-Index Models under Negatively Associated Errors

Xin Qi and Zhuo Xi Yu

1Guangdong Polytechnic of Science and Technology, Zhuhai 519000, China
2School of Economics, Liaoning University, Shenyang 110136, China

Correspondence should be addressed to Xin Qi; xinerqi@sina.com

Received 5 October 2020; Revised 23 February 2021; Accepted 27 February 2021; Published 24 March 2021

Academic Editor: Ding-Xuan Zhou

Copyright © 2021 Xin Qi and Zhuo Xi Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the authors consider the application of the blockwise empirical likelihood method to the partially linear single-index model when the errors are negatively associated, which often exist in sequentially collected economic data. Hereafter, the blockwise empirical likelihood ratio statistic for the parameters of interest is proved to be asymptotically chi-squared. Hence, it can be directly used to construct confidence regions for the parameters of interest. A few simulation experiments are used to illustrate our proposed method.

1. Introduction

The partially linear single-index model is as follows:

\[ Y = g(Z^T \beta) + X^T \theta + \epsilon, \]  

where \( Y \) is a response variable, \( (Z, X) \in \mathbb{R}^p \times \mathbb{R}^q \) is the covariate, \( g(\cdot) \) is an unknown univariate measurable function, \( \epsilon \) is a random error, and \((\beta, \theta)\) is an unknown vector in \( \mathbb{R}^p \times \mathbb{R}^q \) with \( \|\beta\| = 1 \) (where \( \| \cdot \| \) denotes the Euclidean norm). The restriction \( \|\beta\| = 1 \) assures identifiability.

Model (1) is flexible enough to include many important statistical models, so it has attracted much attention and has been extensively studied in recent years. Relevant studies about Model (1) have been done by [1–5]; all of which are based on the independent error sequence. In practice, all the parts of the random error sequence are often associated with each other, such as negatively associated errors, m-dependent errors, and ARCH errors, so that the abovementioned findings cannot be used directly. Therefore, it is necessary to study Model (1) with associated errors.

The finite random variable sequence \( (\xi_i, 1 \leq i \leq n) \) is negatively associated (NA); if, on the condition of any two arbitrary disjoint subsets \( A, B \subseteq \{1, 2, \ldots, n\} \) and any real-valued coordinate-wise nondecreasing functions \( f_1 \) and \( f_2 \), it is established that there is

\[ \text{Cov}\{f_1(\xi_i, i \in A), f_2(\xi_j, j \in B)\} \leq 0. \]

As for infinite random variable sequence, if any arbitrary finite subset is negatively associated, the infinite sequence is negatively associated.

NA sequence has been introduced and studied by the authors of [6, 7] since the 1980s. Because the NA sequence includes the independent sequence, it has been widely applied in multivariate statistical analysis, the permeability analysis, and reliability theory drew much attention, and a lot of research results have been obtained. Under the fields of NA random variables, the author of [8] presented the asymptotic normality and central limit theorem; the authors of [9] proved the law of the iterated logarithm; the authors of [10] studied the exponential inequality, and so on.

However, there is little research about the partially linear single-index model under NA error. This paper, with the enlightenment of [11, 12], focuses on estimating \( \beta, \theta \) with blockwise empirical likelihood when the errors are subjected to NA in Model (1). Throughout this paper, we assume that
the data \(\{X_i, Z_i, Y_i, i = 1, 2, \ldots, n\}\) are generated from Model (1), and \(\{e_i, i = 1, \ldots, n\}\) are NA errors with \(E(e_i|Z_i, X_i) = 0\) and \(\text{Var}(e_i|Z_i, X_i) = \sigma^2_i < \infty\) in Model (1).

The rest of this paper is organized as follows. In Section 2, the blockwise empirical likelihood method and the relative asymptotic result are presented. In Section 3, some simulations are conducted to illustrate the proposed approach. All proofs are shown in Section 4.

2. Methodology and Main Results

2.1. Bias-Corrected Blockwise Empirical Likelihood. In this part, we will use the bias-corrected blockwise empirical likelihood to construct the confidence region for \((\beta, \theta)\). For this reason, first we introduce an auxiliary vector using the bias-corrected method of [2]. The details are as follows. Now that \(\|\beta\| = 1\) means that the value \(\beta\) is a boundary point on the unit sphere and \(g(Z^T\beta)\) does not have a derivative at the point \(\beta\) of the parameter space. Nevertheless, the derivative of \(g(Z^T\beta)\) on \(\beta\) in the building of the empirical likelihood ratio statistics must be adopted. We hereby adopt the "delete-one-component" method which is widely used in the semiparameter model. Let \(\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T\) and \(\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^T\) be a \((p - 1)\)-dimensional parameter vector after deleting the rth component of \(\beta\). Without losing its generality, we assume \(\beta_r > 0\). Then, we can write

\[
l_0(\beta^{(r)}, \theta) = -2 \max \left\{ \sum_{i=1}^n \log(n \pi_i); \; p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \eta_i(\beta^{(r)}, \theta) = 0 \right\}.
\]

Since \(g(\cdot), g'(\cdot), \mu_1(\cdot), \) and \(\mu_2(\cdot)\) are unknown, formula (5) cannot be used to construct the confidence regions directly. To replace them with their estimators is what the researchers usually do. Next, we apply the local linear smooth method of [14] to obtain the estimators of \(g(\cdot)\) and \(g'(\cdot)\). For any fixed \((\beta, \theta)\), we focus on finding a and b to minimize

\[
\sum_{i=1}^n \left[ Y_i - X_i^T \theta - a - b \left( Z_i^T \beta - t \right) \right]^2 K_h \left( Z_i^T \beta - t \right),
\]

where \(K_h(\cdot) = K(-h)/h, K(\cdot)\) is a kernel function and \(h = h(n)\) is a bandwidth. Let \((\bar{a}, \bar{b})\) be the solution to minimize (7). Through a simple calculation,

\[
\bar{g}(t; \beta, \theta) = \bar{a} = \sum_{i=1}^n W_{mi}(t; \beta)(Y_i - X_i^T \theta),
\]

\[
\bar{g}'(t; \beta, \theta) = \bar{b} = \sum_{i=1}^n W_{mi}(t; \beta)(Y_i - X_i^T \theta),
\]

where

\[
W_{mi}(t; \beta) = \frac{n^{-1} K_h \left( Z_i^T \beta - t \right) \left[ S_{n2}(t; \beta) - (Z_i^T \beta - t) S_{n1}(t; \beta) \right]}{S_{n0}(t; \beta) S_{n2}(t; \beta) - S_{n1}^2(t; \beta)},
\]

\[
\bar{W}_{mi}(t; \beta) = \frac{n^{-1} K_h \left( Z_i^T \beta - t \right) \left[ (Z_i^T \beta - t) S_{n0}(t; \beta) - S_{n1}(t; \beta) \right]}{S_{n0}(t; \beta) S_{n2}(t; \beta) - S_{n1}^2(t; \beta)}, \quad i = 1, 2, \ldots, n,
\]

\[
S_{n2}(t; \beta) = \frac{1}{n} \sum_{i=1}^n K_h \left( Z_i^T \beta - t \right) \left( Z_i^T \beta - t \right), \quad l = 0, 1, 2.
\]
Then, we define
\[ \tilde{p}_1(t; \beta) = \sum_{i=1}^{n} W_n(t; \beta)Z_i, \]
and
\[ \tilde{p}_2(t; \beta) = \sum_{i=1}^{n} W_n(t; \beta)X_i, \]
where
\[ \eta_i = Y_i - \tilde{g}(Z_i^T \beta; \theta) - X_i^T \theta \]
as the estimators of \( \mu_1(t) \) and \( \mu_2(t) \), respectively.

When (8), (9), (11), and (12) are plugged in (5) and (6), an estimated auxiliary vector and an estimated bias-corrected empirical log-likelihood ratio can be, respectively, defined as follows:

Under the independent identically distributed errors, the empirical likelihood ratio statistic is constructed by [2]; and its asymptotic result is presented. In this paper, empirical likelihood ratio statistic is constructed by [2]; and the asymptotics of the empirical likelihood ratio statistic is given by [2].

For this, we apply the small-block and large-block arguments to construct the blockwise empirical likelihood ratio. Let \( k = k_n = \lfloor n/(d + s) \rfloor \), where \( \lfloor a \rfloor \) denotes the integral part of \( a \), and \( d = d(n) \) and \( s = s(n) \) are positive integers satisfying \( d + s \leq n \). For \( m = 1, \ldots, k \), let
\[ \omega_{n,2m-1}(\beta^{(r)}, \theta) = e_{n,m}, \]
\[ \omega_{n,2m}(\beta^{(r)}, \theta) = e_{n,m}, \]
\[ \omega_{n,2k+1}(\beta^{(r)}, \theta) = e_{n,k+1}, \]
where
\[ e_{n,m} = \sum_{i=m}^{r_n+d-1} \tilde{\eta}_i(\beta^{(r)}, \theta), \]
\[ e_{n,m} = \sum_{i=m}^{r_n+d-1} \tilde{\eta}_i(\beta^{(r)}, \theta), \]
\[ e_{n,k+1} = \sum_{i=k(d+s)+1}^{n} \tilde{\eta}_i(\beta^{(r)}, \theta), \]
\[ r_n = (m - 1)(d + s) + 1, \]
\[ l_m = (m - 1)(d + s) + d + 1. \]

By using the Lagrange multiplier method, the bias-corrected blockwise empirical likelihood ratio statistic is
\[ \tilde{I}_n(\beta^{(r)}, \theta) = 2 \sum_{j=1}^{2k+1} \log \left( 1 + \lambda^T(\beta^{(r)}, \theta) \omega_{n,j}(\beta^{(r)}, \theta) \right), \]
where \( \lambda \in \mathbb{R}^{p+q-1} \) is determined by
\[ \frac{1}{2k+1} \sum_{j=1}^{2k+1} \frac{\omega_{n,j}(\beta^{(r)}, \theta)}{1 + \lambda^T \omega_{n,j}(\beta^{(r)}, \theta)} = 0. \]

2.2. Asymptotic Result. In this subsection, the main result of this paper is summarized. In order to state the asymptotic result, the following assumptions will be used:

- (C1): the density function \( f(t) \) of \( Z^T \beta \) is bounded away from 0 on \( \tau \) and satisfies the Lipschitz condition of order 1 on \( \tau \), where \( \tau = \{ t = Z^T \beta: Z \in A \} \) and \( A \) is a compact support set of \( Z \).
- (C2): \( g(\cdot), \mu_1(\cdot), \) and \( \mu_2(\cdot) \) have two bounded and continuous derivatives on \( \tau \), where \( \mu_1(\cdot) \) and \( \mu_2(\cdot) \) are the sth and lth components of \( \mu_1(\cdot) \) and \( \mu_2(\cdot) \), respectively.
- (C3): the kernel \( K(\cdot) \) is a bounded symmetric density function and satisfies
\[ \int_{-\infty}^{\infty} u^2 K(u)du \neq 0, \]
\[ \int_{-\infty}^{\infty} |u| K(u)du < \infty, \quad i = 1, 2, \ldots \]

- (C4): \( \sup_{x \in \mathcal{X}} E(x^4|Z = Z, X = x) < \infty; \) \( \sup_{z \in \mathcal{Z}} E(x^4|Z = z) < \infty \)

- (C5): the bandwidth \( h \) satisfies that \( h \to 0, \)
\[ nh^3 \to \infty, \text{ and } nh^8 \to \infty \]

- (C6): \( V(\beta^{(r)}, \theta) = E \{ \Lambda \Lambda^T \} \) and \( \Sigma_e = \text{Cov} \) \( (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \) are both positively definite matrices, where \( \Lambda = [g_t(Z^T \beta)(Z - \mu_t(Z^T \beta))^TF_{\beta^{(r)}}, (X - \mu_2(Z^T \beta))^TF_{\beta^{(r)}}, (X - \mu_2(Z^T \beta))^TF_{\beta^{(r)}}, \ldots, (X - \mu_2(Z^T \beta))^TF_{\beta^{(r)}}] \)

- (C7): \( \sup_{x \in \mathcal{X}} E(\|X\|^2|Z^T \beta = t) < \infty \)
Remark 1. According to [2], (C₁)–(C₃) guarantee the asymptotic distribution theory. (C₅) is a common assumption in the NA situation. (C₆) constrains the block size in order to obtain the desired results.

Theorem 1. Assume that (C₁)–(C₆) are satisfied. If \( \left( \beta^{(r)}, \theta \right) \) is the true value of the parameter and \( \beta_r > 0 \), when \( n \rightarrow \infty \), then
\[
\hat{I}_n(\beta^{(r)}, \theta) \xrightarrow{d} \chi^2_{p+q-1},
\]
where \( \xrightarrow{d} \) stands for the convergence in distribution.

Based on Theorem 1, \( \hat{I}_n(\beta^{(r)}, \theta) \) can be used to construct confidence regions for \( \left( \beta^{(r)}, \theta \right) \). For any given \( 0 < \alpha < 1 \), there exists \( c_\alpha \), which makes \( P\left( X_{p+q-1} > c_\alpha \right) = \alpha \), and then
\[
I_\alpha(\beta^{(r)}, \theta) = \left\{ \left( \beta^{(r)}, \theta \right) \in \hat{I}_n(\beta^{(r)}, \theta) \mid c_\alpha \left\| \beta^{(r)} \right\| < 1 \right\},
\]
which is the confidence regions of \( \left( \beta^{(r)}, \theta \right) \) with the asymptotically correct coverage probability \( 1 - \alpha \).

3. Simulation

In this section, we use two simulation studies to compare the performance of the proposed empirical likelihood method (ELM) and the normal approximation method (NAM).

We assume \( e_i \sim N\left(0, 0.001 i, 1\right), i = 1, \ldots, n \), and \( e_i = e_{i+1} - e_i \). Then, we can get \( e_i \sim N\left(0, 0.001, 2\right) \) for \( 1 \leq i \leq n \) and \( \text{Cov}\left(e_i, e_j\right) = -1 \) for \( |i - j| = 1 \), otherwise \( \text{Cov}\left(e_i, e_j\right) = 0 \). Thus, \( \{e_i\}, 1 \leq i \leq n \) is a NA sequence, and its mean is approximately zero. According to (C₆), the rate of \( h \) is between \( n^{-\left(1/4\right)} \) and \( n^{-\left(1/2\right)} \). Therefore, the bandwidth is selected as \( h = h_{opt} \) by using the cross-validation (CV) method. The details of the CV method have been discussed in the references [15] of the study by Härdle et al. [15]. This selected bandwidth \( h \) satisfies the condition (C₆). Let \( d = \left[n^{\left(1/4\right)}\right] \) and \( s = \left[n^{\left(1/2\right)}\right] \), which satisfy the condition (C₆).

Example 1. In the simulation, we generate 1000 datasets, each consisting of \( n = 100, 200, 300, 500 \). The set of data is generated from the following model:
\[
Y_i = \sin\left(2\pi \cdot Z_i^T \beta\right) + X_i \theta + e_i, \quad 1 \leq i \leq n,
\]
where \( \beta = (1/\sqrt{2}) (1, 1)^T, \theta = 0.5, X_i \sim N(0, 1) \), and \( Z_i \) is the bivariate with independent \( U(0, 1) \) components. The kernel function is taken as the Quartic kernel \( K(t) = \left(15/16\right)(1 - t^2)^2 \) if \( |t| \leq 1 \).

Since \( \beta_1 = \beta_2 \), we only consider the confidence region of the parameter \( (\beta_1, \theta) \). The coverage probabilities of the empirical likelihood confidence regions and the normal approximation confidence regions, with the normal level 0.95, are reported in Table 1. As is expected, the results fit our theory fairly well. The larger the sample size is, the closer the empirical coverage probability is to the nominal level. The proposed empirical likelihood method outperforms the normal approximation method.

Figure 1 plots the proposed empirical likelihood confidence region and the normal approximation confidence regions for \( (\beta_1, \theta) \) based on the confidence level of 0.95 when the sample size is 300.

Example 2. In this simulation, the coverage probabilities and average lengths of confidence intervals are calculated by the proposed empirical likelihood method and the normal approximation method. Consider Model (1) with \( p = 3 \) and \( q = 2 \), where \( \beta = (1/\sqrt{4}) (1, 2, 3)^T, \theta = (0.2, 0.7)^T \), and \( g() = 1.5 \sin() \). \( Z \) is independent and all from the uniform \( U(0, 1) \), and the two components of \( X \) are from the bivariate standard normal distribution. The kernel function is taken as the Epanechnikov kernel \( K(t) = (3/4)(1 - t^2)^2 \) if \( |t| \leq 1 \).

Based on 500 simulation runs, the simulation results are reported in Table 2. From Table 2, the following results can be obtained. The coverage probabilities of the empirical likelihood method and the normal approximation method are in agreement with the nominal level of 0.90; the empirical likelihood method has slightly smaller interval lengths compared with the normal approximation method.

4. Proofs

In order to prove Theorem 1, we first give some lemmas. Throughout this section, for a concise and convenient representation, we use \( c(0 < c < \infty) \) to denote any constant which may take a different values for each appearance, use \( \lambda_{min}(A) \) and \( \lambda_{max}(A) \) to denote the smallest and largest eigenvalues of \( A \), respectively, and write \( A^{1/2} = A \cdot A^T \).

Lemma 1. Let \( \{a_i, i = 1, 2, \ldots, n\} \) and \( \{b_i, i = 1, 2, \ldots, n\} \) be any two sequences, then
\[
\sum_{i=1}^{n} a_i b_i \leq c \cdot \max_{1 \leq s \leq n} |b_s| \cdot \max_{1 \leq k \leq n} \sum_{i=1}^{k} a_i,
\]
where \( \{j_1, j_2, \ldots, j_n\} \) is an arbitrary arrangement of order \( n \).

The proof of Lemma 1. Firstly, we assume \( b_i \geq 0, i = 1, 2, \ldots, n \). Via Abel inequality, it follows that
\[
\sum_{i=1}^{n} a_i b_i \leq 2 \max_{1 \leq s \leq n} |b_s| \cdot \max_{1 \leq k \leq n} \sum_{i=1}^{k} a_i.
\]

Secondly, we assume \( b_i < 0, i = 1, 2, \ldots, n \), and it also follows that
\[
\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} (a_i \cdot b_i) \leq 2 \max_{1 \leq s \leq n} |b_s| \cdot \max_{1 \leq k \leq n} \sum_{i=1}^{k} a_i.
\]

Combining (23) and (24), then we get
Lemma 2. Let $a_1, a_2, \ldots, a_n$ be any random variables with $\max_{1 \leq i \leq n} E|a_i|^s \leq c < \infty$ for some constants $s > 0$ and $c > 0$. Then,

$$\max_{1 \leq i \leq n} |a_i| = O_p\left(n^{1/s}\right).$$

(26)

Refer to the proof of Lemma 1 of [11].

The proof of Lemma 3 can be finished with the work by [16].

Lemma 3. Let $\{\xi_j; 1 \leq j \leq n\}$ be NA random variables with $E|\xi_j| = 0, E|\xi_j|^s < \infty (s > 1)$ and $\{a_j; j \geq 1\}$ be a sequence of real constants. Then, there exists a constant $c > 0$ which only depends on the given number $s$ so that

$$\max_{1 \leq j \leq n} |\xi_j| = O_p\left(n^{2/s}\right).$$

(27)

$\max_{1 \leq j \leq n} |\xi_j| = O_p\left(n^{2/s}\right).$

(27)

Refer to the proof of Lemma 1 of [11].

The proof of Lemma 3 can be finished with the work by [16].

Lemma 4. Assume that $\{C_r\}$ hold, for any integer $2 \leq r \leq 4$, we can obtain, uniformly over $1 \leq i \leq n$,
where \( \hat{\varphi}() = \tilde{g}(\cdot), \tilde{\mu}_{1l}(\cdot), \tilde{\mu}_{2l}(\cdot) \) and \( \varphi(\cdot) = g(\cdot), \mu_{1l}(\cdot), \mu_{2l}(\cdot) \). Here, \( \mu_{1l}(\cdot) \) and \( \mu_{2l}(\cdot) \) are, respectively, the \( l \)th component of \( \mu_{1}(\cdot) \) and \( \mu_{2}(\cdot) \), \( 1 \leq s \leq p, q \leq l \leq q \).

The proof of Lemma 4 is similar to the proof of Lemma 3 by [12]. So, the details are omitted here.

**Lemma 5.** Suppose that it satisfies (C₁)–(C₅), \( (\beta^{(r)}, \theta) \) is the true value of the parameter, and the \( r \)th component of \( \beta > 0 \), then

\[
\begin{align*}
E\left\{ \tilde{\varphi}(Z_{i}^T \beta; \theta) - \varphi(Z_{i}^T \hat{\beta}) \right\} &= O(h^2) + O(n^{-(r/2)}h^{1-r}), \\
E\left\{ \tilde{g}'(Z_{i}^T \beta; \theta) - g'(Z_{i}^T \hat{\beta}) \right\} &= O(h^2) + O(n^{-(r/2)}h^{1-2r}),
\end{align*}
\]

(28)

\[
C = \left( \begin{array}{c} C_1 \end{array} \right)
\]

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}(\beta^{(r)}, \theta) \rightarrow N(0, \Gamma),
\]

(29)

and

\[
\frac{1}{n} \sum_{i=1}^{n} \omega_{i} \bar{\omega}_{i}^T \Gamma = o_p(1),
\]

(30)

where \( \Gamma = E[(\tilde{\Delta}e)^{\otimes 2}] \), and \( \tilde{\Delta} \) is defined in (C₆).

The proof of Lemma 5. It is easy to show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\eta}(\beta^{(r)}, \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\Delta}_{i} \varepsilon_{i} + \sum_{i=1}^{\tilde{r}} M_{i},
\]

(31)

where

\[
\begin{align*}
M_{11} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \mu_{1}(Z_{i}^T \beta) - \tilde{\mu}_{1}(Z_{i}^T \hat{\beta}) \right] g'(Z_{i}^T \hat{\beta}) \varepsilon_{i}, \\
M_{12} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \mu_{2}(Z_{i}^T \beta) - \tilde{\mu}_{2}(Z_{i}^T \hat{\beta}) \right] \varepsilon_{i}, \\
M_{21} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \tilde{g}'(Z_{i}^T \beta; \theta) - g'(Z_{i}^T \hat{\beta}) \right] \cdot \left[ Z_{i} - \tilde{\mu}_{1}(Z_{i}^T \hat{\beta}; \theta) \right] \varepsilon_{i}, \\
M_{31} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(Z_{i}^T \beta) - \tilde{g}(Z_{i}^T \hat{\beta}; \theta) \right] \cdot \left[ Z_{i} - \tilde{\mu}_{1}(Z_{i}^T \hat{\beta}; \theta) \right] g'(Z_{i}^T \hat{\beta}), \\
M_{32} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(Z_{i}^T \beta) - \tilde{g}(Z_{i}^T \hat{\beta}; \theta) \right] \cdot \left[ Z_{i} - \tilde{\mu}_{1}(Z_{i}^T \hat{\beta}; \theta) \right], \\
M_{41} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g(Z_{i}^T \beta) - \tilde{g}(Z_{i}^T \hat{\beta}; \theta) \right] \left[ \tilde{g}'(Z_{i}^T \beta; \theta) - g'(Z_{i}^T \hat{\beta}) \right] \cdot \left[ Z_{i} - \tilde{\mu}_{1}(Z_{i}^T \hat{\beta}; \theta) \right].
\end{align*}
\]

In order to prove (29), we firstly need to show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\Delta}_{i} \varepsilon_{i} \rightarrow N(0, \Gamma),
\]

(33)

Let \( \Delta_{n} = (1/n) \sum_{i=1}^{n} \tilde{\Delta}_{i} \varepsilon_{i} \). For the central limit theorem of \( (n\Delta_{n})^{(1/2)} \sum_{i=1}^{n} \tilde{\Delta}_{i} \varepsilon_{i} \), we refer to Lemma 5 of [11]; then, we can obtain (33).
Secondly, we need to show that $M_{m1} \rightarrow^p 0$ ($u = 1, 2, 3, 4$) and $M_{m2} \rightarrow^p 0$ ($m = 1, 3$). Here, we consider $M_{12}$. Let $M_{12,s}$ denote the $st$ component of $M_{12}$. By (C2) and (C3) and Lemma 1 and Lemma 4, we can get

$$E(M_{12,s}) \leq c \cdot \sup_{x} E(\epsilon_i^2|Z_i = x) \sum_{j=1}^{m} E(\mu_2(Z_i^T \beta) - \tilde{\mu}_2(Z_i^T \tilde{\beta})^2) \leq ch^4 + c(nh)^{-1} \rightarrow 0. \quad (34)$$

The other formulas mentioned above can be proved by using Lemma 1, Lemma 3, and Lemma 4 and the similar methods of [2]. These formulas, together with (31) and (33), complete the proof of (29).

The proof of (30) uses a similar method to the proof of Lemma 5 by [12]. Here, we only give some key steps. For any $c \in \mathbb{R}^{P \times q}$ with $\|c\| = 1$, we will show that

$$1 \over n c^T T_m n c = 1 \over n \sum_{i=1}^{k} \left( c^T \Gamma^{(r/2)} \sum_{j=r}^{r+d-1} \bar{\lambda}_j \epsilon_j + c^T \Gamma^{(r/2)} \sum_{j=r}^{r+d-1} R_{n j} \right)^2 = 1 \over n \sum_{i=1}^{k} (\nu_i + \nu_{i2})^2, \quad (39)$$

where

$$R_{n j} = \left( \left( \epsilon_j g((Z_j^T \tilde{\beta} - \bar{\mu}_j)(Z_j^T I_{K^{(r)}}, \epsilon_j \tilde{\beta}_{i j}) - (\epsilon_j g((Z_j^T \tilde{\beta} - \bar{\mu}_j), (Z_j^T I_{K^{(r)}}, 0)) \right)^T \right)^T + \left( \left( \bar{\gamma}_j g((Z_j^T \tilde{\beta} - \bar{\mu}_j)^T Z_j^T I_{K^{(r)}}, \bar{\gamma}_j \tilde{X}_j^T - (\bar{\gamma}_j g((Z_j^T \tilde{\beta} - \bar{\mu}_j), 0)^T, \right), \bar{\mu}_{i j} = \mu_i (Z_j^T \tilde{\beta}) - \bar{\mu}_i (Z_j^T \tilde{\beta}), \bar{\mu}_{i j} = \mu_i (Z_j^T \tilde{\beta}) - \bar{\mu}_i (Z_j^T \tilde{\beta}), \tilde{Z}_j = Z_j - \bar{\mu}_i (Z_j^T \tilde{\beta}), \bar{X}_j = X_j - \bar{\mu}_i (Z_j^T \tilde{\beta}), \bar{\gamma}_j = g(Z_j^T \tilde{\beta}) - g((Z_j^T \tilde{\beta}), 0)^T, \bar{\gamma}_j = g(Z_j^T \tilde{\beta}) - g((Z_j^T \tilde{\beta}), 0)^T. \quad (40)$$

As arguments in [11], we can get (36). Similar decomposition and proof can be used in (37) and (38). Thus, (30) is completely proved.

**Lemma 6.** Under (C1)–(C4), it follows that

$$\max_{1 \leq i \leq 2k} \|w_i\| = o_p(n^{1/2}). \quad (41)$$

The proof of Lemma 6. Now that

$$\max_{1 \leq i \leq 2k} \|w_i\| \leq c \cdot \max_{1 \leq i \leq 2k} \|\tilde{w}_i (\tilde{\beta}^{(r)}, \theta)\|, \quad (42)$$

we only need to prove

$$\max_{1 \leq i \leq 2k} \|\tilde{w}_i (\tilde{\beta}^{(r)}, \theta)\| = o_p(n^{1/2}). \quad (43)$$

Let $\tilde{w}_i$ denote the $st$ component of $\tilde{w}_i$ and the notation of Lemma 5 is used.

We consider the later $q$ components of $\tilde{w}_i (\tilde{\beta}^{(r)}, \theta)$. By a simple calculation, we can obtain

$$\max_{1 \leq i \leq 2k} \|\tilde{w}_i (\tilde{\beta}^{(r)}, \theta)\| \leq c \cdot \max_{1 \leq i \leq 2k} |\bar{\lambda}_{i,s} \epsilon_i + c \cdot \max_{1 \leq i \leq 2k} |\bar{\mu}_{i,s} \epsilon_i| + c \cdot \max_{1 \leq i \leq 2k} |\bar{\gamma}_j \bar{X}_{j,i,s}| \quad (44)$$

Applying Lemma 2 with $E(\epsilon_i^2) < \infty$,

$$\max_{1 \leq i \leq 2k} |\epsilon_i| = O_p(n^{1/2}). \quad (45)$$

This implies that $A_{n,s} = o_p(n^{1/2})$ since $E(|\bar{\lambda}_{i,s} \epsilon_i|^2) < \infty$. Combining the Markov inequality, Lemma 4 and (C4), for any $\delta > 0$,
\[ P(A_{n,2} > \delta \sqrt{n}) \leq \sum_{i=1}^{n} P\left( \left| \mu_n (Z_i^T \hat{\beta}) - \tilde{\mu}_n (Z_i^T \hat{\beta}; \hat{\beta}) \right| > \delta \sqrt{n} \right) \]
\[
\leq \frac{1}{n \delta^2} \sup_{z} E(\varepsilon^2 | Z = z) \sum_{i=1}^{n} E\left( \left| \mu_n (Z_i^T \hat{\beta}) - \tilde{\mu}_n (Z_i^T \hat{\beta}; \hat{\beta}) \right|^2 \right),
\]
\[
\leq c h^4 + c (nh)^{-1} \rightarrow 0.
\]

Hence, \( A_{n,2} = o_p(n^{(1/2)}) \). \( A_{n,3} \) can be dealt similarly. Then,
\[
\max_{1 \leq i \leq n} |\tilde{\eta}_n (\hat{\beta}^{(r)}, \theta) | = o_p(n^{(1/2)}), \quad \text{for } p \leq s \leq p + q - 1.
\]

Moreover,
\[
\max_{1 \leq i \leq n} |\tilde{\eta}_n (\hat{\beta}^{(r)}, \theta) | = o_p(n^{(1/2)}), \quad \text{for } 1 \leq s \leq p - 1,
\]
\[\text{(47)}\]
\[\text{and } (48)\]
can be similarly dealt with. Thus, (43) can be shown.

\[
0 = \frac{1}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}} = \frac{1}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}} - \frac{\rho}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)} - \frac{\rho}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)} = \frac{\rho}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)} - \frac{\rho}{2k + 1} \mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)}.
\]
\[\text{(50)}\]

It follows that
\[
|\mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}} | \geq \rho \lambda_{\text{min}} (S_1) - \rho |\mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)}| \geq \rho \lambda_{\text{min}} (S_1) - \rho |\mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)}|.
\]
\[\text{(51)}\]

\[\text{By a straightforward calculation, we can get} \]
\[
\|\mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)} \| \geq \rho \lambda_{\text{min}} (S_1) - \rho |\mathbf{\tilde{\omega}}^T \mathbf{\tilde{\omega}}^{(1/2)} \mathbf{\tilde{\omega}}^{(1/2)}|.
\]
\[\text{(52)}\]

As discussed in \[12\], we obtain
\[
\Gamma = o_p(n).
\]
\[\text{(53)}\]

Using Lemma 5, Lemma 6, and (53), we can have
\[
\rho \left( 1 - \max_{1 \leq j \leq 2k + 1} \left\| w_n \right\| \right) o_p(n^{(-1/2)}) \leq o_p(n^{(-1/2)}).
\]
\[\text{(54)}\]

Therefore, it is easy to obtain
\[
\rho = \| \lambda (\hat{\beta}^{(r)}, \theta) \| = O_p(n^{(-1/2)}).
\]
\[\text{(55)}\]
\[ \phi_1 \leq \max_{j=1,2k+1} \| y_j \|^3 (2k+1)\max_{j=1,2k+1} \| w_{nj} \|_2 \left( \frac{1}{2k+1} \sum_{j=1}^{2k+1} w_{nj} \omega_n_j^T \right) = o_p(1). \] (57)

Thus,
\[ \bar{\ell}_n(\beta^{(r)}, \theta) = 2(2k+1)\lambda^T \bar{w} - (2k+1)\lambda^T S_1 \lambda + o_p(1). \] (58)

By calculating directly from (17), we obtain
\[ \phi_2 = \sum_{j=1}^{2k+1} \lambda_j \omega_n_j y_j^2 \leq \lambda^T \Gamma^{-(1/2)} \max_{j=1,2k+1} \| w_{nj} \|_2 \sum_{j=1}^{2k+1} \omega_n_j \omega_n_j^T = o_p(n^{-1/2}). \] (60)

Therefore, it follows that
\[ (2k+1)(\Gamma^{-(1/2)} S_1 \Gamma^{-(1/2)}) \Gamma^{-(1/2)} \lambda = (2k+1)\Gamma^{-(1/2)} \bar{w} + o_p(n^{-1/2}). \] (61)

It is easy to get
\[ \bar{\ell}_n(\beta^{(r)}, \theta) = (2k+1)\left( \frac{\Gamma^{-(1/2)} \bar{w}}{\Gamma^{-(1/2)} S_1 \Gamma^{-(1/2)}} \right)^T \frac{\Gamma^{-(1/2)} \bar{w}}{\Gamma^{-(1/2)} S_1 \Gamma^{-(1/2)}} + o_p(1) \] (63)
\[ = \left( (2k+1)\Gamma^{-(1/2)} \bar{w} \right)^T \left( (2k+1)\Gamma^{-(1/2)} S_1 \Gamma^{-(1/2)} \right)^{-1} \left( (2k+1)\Gamma^{-(1/2)} \bar{w} \right) + o_p(1). \]

Applying Lemma 5, we have
\[ \left( (2k+1)\Gamma^{-(1/2)} \bar{w} \right)^T \left( (2k+1)\Gamma^{-(1/2)} S_1 \Gamma^{-(1/2)} \right)^{-1} \left( (2k+1)\Gamma^{-(1/2)} \bar{w} \right) \xrightarrow{L} \chi_{p.q-1}^2. \] (64)

Thus, we complete the proof of Theorem 1.

**Data Availability**

Simulation data were obtained from Monte Carlo experiment.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Authors’ Contributions**

All authors read and approved the final manuscript.

**Acknowledgments**

This work was supported by the Philosophy and Social Sciences Planning Project of Guangdong Province during the “13th Five-Year” Plan Period (nos. GD18CYJ08 and GD20CJY50), National Social Science Foundation of China (no. 18CQT032), Guangdong Province Educational Science “Thirteenth Five-Year Plan” Project (no. 2019GJXJ272), Guangdong Province Research Project (no. 2020WT030), Guangdong Provincial Department of Education Project (no. 2020WQNCX141), and Guangdong Polytechnic of Science and Technology Research Project (nos. XJJPY2018006 and JG201918).

**References**


