



Article

Robust Estimation and Tests for Parameters of Some Nonlinear Regression Models

Pengfei Liu ^{1,2,†} , Mengchen Zhang ^{3,*,†} , Ru Zhang ^{1,4,*,†} and Qin Zhou ^{1,2,†}

¹ School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, China; liupengfei@jsnu.edu.cn (P.L.); graceqinzhou@jsnu.edu.cn (Q.Z.)

² Research Institute of Mathematical Sciences, Jiangsu Normal University, Xuzhou 221116, China

³ Department of Public Administration and Policy, College of Public Affairs, National Taipei University, Taipei 237, Taiwan

⁴ Jiangsu Provincial Key Laboratory of Educational Big Data Science and Engineering, Jiangsu Normal University, Xuzhou 221116, China

* Correspondence: s810671402@webmail.ntpu.edu.tw (M.Z.); 2020180327@jsnu.edu.cn (R.Z.)

† The author order is alphabetically sorted. These authors contributed equally to this work. And the submit author is Ru Zhang.

Abstract: This paper uses the median-of-means (MOM) method to estimate the parameters of the nonlinear regression models and proves the consistency and asymptotic normality of the MOM estimator. Especially when there are outliers, the MOM estimator is more robust than nonlinear least squares (NLS) estimator and empirical likelihood (EL) estimator. On this basis, we propose hypothesis testing Statistics for the parameters of the nonlinear regression models using empirical likelihood method, and the simulation performance shows the superiority of MOM estimator. We apply the MOM method to analyze the top 50 data of GDP of China in 2019. The result shows that MOM method is more feasible than NLS estimator and EL estimator.



Citation: Liu, P.; Zhang, M.; Zhang, R.; Zhou, Q. Robust Estimation and Tests for Parameters of Some Nonlinear Regression Models.

Mathematics **2021**, *9*, 599. <https://doi.org/10.3390/math9060599>

Academic Editor: Vasile Preda

Received: 15 February 2021

Accepted: 5 March 2021

Published: 11 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: median-of-means (MOM); nonlinear regression (NR); empirical likelihood (EL); hypothesis testing (HT)

1. Introduction

A nonlinear regression model refers to a regression model in which the relationship between variables is not linear. Nonlinear regression model has been widely used in various disciplines. For instance, Hong [1] applied a nonlinear regression model to the economic system prediction; Wang et al. [2] studied the application of nonlinear regression model in the detection of protein layer thickness; Chen et al. [3] utilized a nonlinear regression model in the price estimation of surface-to-air missiles; Archontoulis and Miguez [4] used a nonlinear regression model in agricultural research.

The principle of median-of-means (MOM) was firstly introduced by Alon, Matias, and Szegedy [5] in order to approximate the frequency moment with space complexity. Lecué and Lerasle [6] proposed new estimators for robust machine learning based on MOM estimators of the mean of real-valued random variables. These estimators achieved optimal rates of convergence under minimal assumptions on the dataset. Lecué et al. [7] proposed MOM minimizers estimator based on MOM method. The MOM minimizers estimator is very effective when the instantaneous hypothesis may have been corrupted by some outliers. Zhang and Liu [8] applied MOM method to estimate the parameters in multiple linear regression models and AR error models of repeated measurement data.

For unknown parameters of a nonlinear regression model, Radchenko [9] proposed an estimator named nonlinear least square to approximate the unknown parameters. Ding [10] introduced the empirical likelihood (EL) estimator of the parameter of the nonlinear regression model based on the empirical likelihood method. However, when there are outliers, the general methods are more sensitive and easily affected by the outliers based

on Gao and Li [11]. On the basis of the study of Zhang and Liu [8], this paper applies the MOM method to estimate the parameters of the nonlinear regression models and receives more robust results.

The paper is organized as follows: In Section 2, we review the definition of the nonlinear regression model and introduce the MOM method specifically. We prove the consistency and asymptotic properties of the MOM estimator. In Section 3, we introduce a new test method based on the empirical likelihood method for the median. Section 4 illustrates the superiority of the MOM method with simulation studies. A real application to GDP data is given in Section 5, and the conclusion is discussed in the last section.

2. Median-of-Means Method Applies to Nonlinear Regression Model

We consider the following nonlinear regression model introduced by Wu [12]

$$y_i = g(\theta, x_i) + \varepsilon_i \quad i = 1, \dots, T. \quad (1)$$

where $\theta = (\theta_1, \dots, \theta_k)^T$ is a fixed $k \times 1$ unknown parameter column vector. x_i is the i -th “fixed” input vector with observation y_i . $g(\theta, x_i)$ is a known functional form (usually nonlinear). ε_i are i.i.d errors with 0 mean and σ^2 unknown variance.

According to Zhang and Liu [8], MOM estimator of θ is produced by the following steps:

Step I: We separate $(y_i, x_i), i = 1, \dots, T$ into g groups. The number of observations in each group is $n = T/g$ (Usually for the convenience of calculation, we assume that T is always divisible by g). We discuss the choice of grouping number g . According to the suggestion by Emilien et al. [13], $g = \lceil 8 \times \log(\frac{1}{\zeta}) \rceil$ for any $\zeta \in (0, 1)$, where $\lceil \eta \rceil$ is the ceiling function. In fact, the structure of observations is always unknown, and the diagnosis of outliers is complicated. Therefore, we usually set $\zeta = \frac{C}{\sqrt{T}}$ for some constant C regardless of outliers.

Step II: We estimate the parameter θ in each group $j, 1 \leq j \leq g$ by the nonlinear least square estimator $\hat{\theta}^{(j)} = (\hat{\theta}_1^{(j)}, \dots, \hat{\theta}_k^{(j)})^T$.

Step III: The MOM estimator of $\hat{\theta}^{MOM} = (\hat{\theta}_1^{MOM}, \dots, \hat{\theta}_k^{MOM})^T$ is defined, where $\hat{\theta}_q^{MOM} = \text{median}(\hat{\theta}_q^{(1)}, \dots, \hat{\theta}_q^{(g)})$, $q = 1, \dots, k$.

The asymptotic properties of $\hat{\theta}^{MOM}$ are summarized in the following theorems. Their proofs are postponed to Appendix A.

Theorem 1. For some constant C and any positive integer g , we suppose the following:

(I) For certain $0 < a < b < \infty$ and any $\theta_1, \theta_2 \in \Theta$, Θ is an open interval (finite or infinite) of the real axis E_1 . $\varphi_n(\theta_1, \theta_2) = \infty$ for $\theta_1 \neq \theta_2$ if at least one of the points θ_1, θ_2 is $-\infty$ or ∞ . $i = 1, \dots, n$.

$$a(\theta_1 - \theta_2)^2 \leq \varphi_n(\theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^n [g(\theta_1, x_i) - g(\theta_2, x_i)]^2 \leq b(\theta_1 - \theta_2)^2$$

Suppose $E|\varepsilon_1|^s < \infty$ for some $s \geq 2$. For $n \geq N_0$ and sufficiently large positive ρ, c does not depend on n and ρ .

(II) $g'(\theta_0, x_i), g''(\theta_0, x_i)$ exist for all θ_0 near $\theta_q, q = 1, \dots, k$, the true value θ_q is in the interior of Θ , and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [(g'(\theta_q, x_i))^2] = S \quad 0 < S < \infty$$

(III) There exists $\theta_v \in \Theta$ as $n \rightarrow \infty$ and $|\theta_v - \theta_q| \rightarrow 0$.

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \{(g'(\theta_v, x_i))^2\}}{\sum_{i=1}^n \{(g'(\theta_q, x_i))^2\}} = 1$$

(IV) There exists a $\delta > 0$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{|\theta_0 - \theta_q| \leq \delta} \left\{ \frac{\partial^2 g(\theta_0, x_i)}{\partial \theta_0^2} \right\}^2 < \infty$$

for all $i = 1, \dots, n$, where $s = \{\theta_q \in \Theta, |\theta_0 - \theta_q| \leq \delta\}$.

According to conditions I \sim IV, for any fixed $x > 0$, we can get

$$\mathbb{P}(|\hat{\theta}_q^{\text{MOM}} - \theta_q| \geq x) \leq \frac{C}{(T/g)^{8/5}}. \quad (2)$$

Theorem 2. (1) Suppose g is fixed and $\sigma \neq 0$. Let $\Theta_1, \Theta_2, \dots, \Theta_g$ be i.i.d standard normal random variables. When $T \rightarrow \infty$,

$$\frac{\sqrt{n}}{\hat{\sigma}_n S^{-1/2}} \xrightarrow{d} \text{median}\{\Theta_1, \dots, \Theta_g\}. \quad (3)$$

(2) Suppose $T/g^2 \rightarrow \infty$ as $g \rightarrow \infty$ and $\sigma \neq 0$. Afterwards the following asymptotic normal holds

$$\frac{\sqrt{T}}{\hat{\sigma}_n S^{-1/2}} \xrightarrow{d} \sqrt{2/\pi} N(0, 1). \quad (4)$$

3. Empirical Likelihood Test Based on MOM Method

In Section 2, this paper uses the MOM method to estimate the parameters of the nonlinear regression model. In this section, we consider the hypothesis test that θ equals a given value parameter based on the empirical likelihood method.

Because different groups are disjoint, $\hat{\theta}_q^{(1)}, \hat{\theta}_q^{(2)}, \dots, \hat{\theta}_q^{(j)}, j = 1, \dots, g, q = 1, \dots, k$ are i.i.d. We treat them as a sample and apply empirical likelihood. For each j , we say $T_{n,j} = I(\hat{\theta}_q^{(j)} \leq \theta_q)$. Obviously, $\mathbb{E}T_{n,j} \approx 0.5$. In fact, $\mathbb{E}T_{n,j} - 0.5 = O(n^{-1/2})$ (by the process of proof of Theorem in the Appendix A). Given restrictive conditions, the empirical likelihood ratio of θ is

$$R(\theta) = \max \left\{ \prod_{j=1}^g g \omega_j \mid \sum_{j=1}^g \omega_j T_{n,j} = 0.5, \omega_j \geq 0, \sum_{j=1}^g \omega_j = 1 \right\}. \quad (5)$$

Using the Lagrange multiplier to find the maximum point we obtained the following equation.

$$\omega_j = \frac{1}{g} \frac{1}{1 + \lambda(T_{n,j} - 0.5)}$$

where $\lambda = \lambda(\theta)$ satisfies the equation

$$0 = \frac{1}{g} \sum_{j=1}^g \frac{T_{n,j} - 0.5}{1 + \lambda(T_{n,j} - 0.5)}. \quad (6)$$

Theorem 3. According to Theorem 2 and Owen [14], as $g, n \rightarrow \infty$, we have

$$-2 \log R(\theta) \xrightarrow{d} \chi_1^2. \quad (7)$$

Using the Theorem 3, the rejection region for the hypothesis with significance level α ($0 < \alpha < 1$)

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

can be constructed as

$$\mathbb{R} := \{-2 \log R(\theta) > \chi_1^2(\alpha)\}$$

where $\chi_1^2(\alpha)$ is the upper α -th quantile of χ_1^2 .

4. Simulation Study

In this section, we use R software for simulation. Simulation experiments are carried out to compare the performance of the MOM estimator with the nonlinear least squares (NLS) estimator and the EL estimator under “no outliers” and “with outliers” cases in Examples 1–3. The definition of Mean Square Error (MSE) of $\hat{\theta}^{EL}$, $\hat{\theta}^{MOM}$ and $\hat{\theta}^{NLS}$ are as follows.

$$MSE = \frac{1}{D-1} \sum_{i=1}^D (\hat{\theta}_q - \theta_q)^2, \quad q = 1, \dots, k. \quad (8)$$

$\hat{\theta}_q, \theta_q$ represent the estimated value of the parameter and the true value of the parameter respectively in formula (8). D represents the total number of simulations, and in this article, $D = 1000$. The MSE results calculated in Tables 1–3 are all multiplied by 100. The results are accurate to three decimal places. In Examples 4–6, we compare our proposed method with empirical likelihood inference proposed by Jiang [15].

We report the empirical sizes and powers of the two methods, where size represents the probability of rejecting the null hypothesis provided it is true. In this paper, we set that the nominal significance level is 0.05. If the value is close to 0.05 it is good. Power represents the probability of rejecting the null hypothesis provided it is false. If the value of power is close to 1 it is good. Empirical size or power represents $\frac{n_1}{D}$, where n_1 refers to the number of times the null hypothesis is rejected in D simulations. In Tables 4–6 of this article, the size value refers to the empirical likelihood, and power refers to the empirical power. In fact, the empirical size is the estimated value of size, and the empirical power is the estimated value of power. We consider the following three forms of nonlinear regression models, which were also considered by Hong [16].

$$\text{model1: } y_i = 0.8^{x_i} + \varepsilon_i, \quad i = 1, \dots, T.$$

$$\text{model2: } y_i = x_i^{0.6} + \varepsilon_i, \quad i = 1, \dots, T.$$

$$\text{model3: } y_i = e^{(0.5x_i)} + \varepsilon_i, \quad i = 1, \dots, T.$$

In this paper, for convenience, we fix the number of groups in simulation. We find that the result is consistent with the calculation result according to the formula $g = \lceil 8 \times \log(\frac{1}{\zeta}) \rceil$ which suggested by Emilien et al. [13]. Throughout the paper, the distribution abbreviations B, U, N, P represent binomial distribution, uniform distribution, normal distribution respectively and Poisson distribution. $N(0,1)$ represents the standard normal distribution. We set the number of repeated observations T to 100, 200, ..., 1000.

Example 1. We consider model $y_i = 0.8^{x_i} + \varepsilon_i$. For the observation data, the grouping is carried out according to the grouping principle. Taking the effect of the measures of dispersion in data sets into consideration (accuracy of the estimator may be affected by the dispersion in the data set). x_i are generated from the $P(0.7)$, ε_i are generated from $N(0,1)$. The output variable y_i has outliers. There are three cases of outliers. We choose 1% T outliers from $B(20, 1/2)$, 2% T outliers from $U(7, 8)$ and 2% T outliers from $N(6, 2)$, respectively. The results are shown in Table 1.

Example 2. We consider model $y_i = x_i^{0.6} + \varepsilon_i$, x_i are generated from the $U(2, 3)$, ε_i are generated from $N(0, 1)$. The output variable y_i have outliers. There are three cases of outliers. We choose 1% T outliers from $B(22, 1/2)$, 2% T outliers from $U(7, 8)$ and 2% T outliers from $N(7, 3)$, respectively. The results are shown in the Table 2.

Table 1. Mean Square Error (MSE) for $\hat{\theta}^{NLS}$, $\hat{\theta}^{MOM}$ and $\hat{\theta}^{EL}$ in Example 1.

T	No Outliers			1% from $B(20, \frac{1}{2})$			2% from $U(7, 8)$			2% from $N(6, 2)$		
	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS
100	1.617	1.375	1.329	2.208	2.020	2.081	2.321	1.969	2.023	2.065	1.790	1.792
200	0.650	0.669	0.649	1.230	1.071	1.229	2.000	1.727	1.753	1.610	1.486	1.611
300	0.415	0.421	0.414	1.011	0.867	1.012	1.798	1.714	1.799	1.157	1.058	1.158
400	0.322	0.328	0.321	0.830	0.717	0.831	1.351	1.178	1.351	1.134	1.035	1.135
500	0.268	0.274	0.267	0.779	0.628	0.780	1.256	1.104	1.256	1.129	1.002	1.128
600	0.212	0.214	0.212	0.729	0.578	0.729	1.145	0.989	1.146	1.036	0.874	1.036
700	0.173	0.174	0.173	0.697	0.554	0.698	1.195	0.979	1.196	1.015	0.848	1.014
800	0.162	0.163	0.161	0.667	0.486	0.668	1.119	0.976	1.120	1.016	0.828	1.017
900	0.141	0.142	0.141	0.637	0.448	0.638	1.113	0.891	1.113	1.007	0.832	1.008
1000	0.128	0.129	0.127	0.622	0.441	0.623	1.083	0.865	1.083	0.996	0.809	0.997

Table 2. MSE for $\hat{\theta}^{NLS}$, $\hat{\theta}^{MOM}$, $\hat{\theta}^{EL}$ in Example 2.

T	No Outliers			1% from $B(22, \frac{1}{2})$			2% from $N(7, 3)$			2% from $U(7, 8)$		
	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS
100	0.380	0.384	0.381	0.626	0.598	0.627	0.822	0.809	0.823	0.606	0.587	0.607
200	0.189	0.193	0.190	0.481	0.455	0.480	0.569	0.554	0.570	0.596	0.590	0.597
300	0.126	0.127	0.126	0.414	0.399	0.415	0.525	0.522	0.526	0.578	0.563	0.577
400	0.100	0.100	0.099	0.379	0.344	0.380	0.472	0.465	0.472	0.547	0.544	0.546
500	0.078	0.079	0.077	0.370	0.328	0.369	0.447	0.428	0.447	0.533	0.515	0.533
600	0.061	0.065	0.063	0.346	0.301	0.347	0.458	0.415	0.459	0.501	0.494	0.501
700	0.055	0.056	0.055	0.337	0.296	0.336	0.424	0.401	0.425	0.498	0.492	0.498
800	0.049	0.049	0.048	0.336	0.278	0.336	0.436	0.403	0.437	0.496	0.472	0.495
900	0.042	0.044	0.042	0.342	0.276	0.341	0.410	0.369	0.410	0.490	0.464	0.489
1000	0.038	0.042	0.038	0.332	0.278	0.333	0.420	0.387	0.420	0.487	0.455	0.487

Example 3. We consider model $y_i = e^{(0.5x_i)} + \varepsilon_i$, x_i are generated from $U(-1, 0)$. ε_i are generated from $N(0, 1)$. The output variable y_i have outliers. There are three cases of outliers. We choose 1%T outliers from $B(20, 1/2)$, 2%T outliers from $N(6, 2)$ and 2%T outliers from $U(6, 7)$, respectively. The results are shown in the Table 3.

Table 3. MSE for $\hat{\theta}^{NLS}$, $\hat{\theta}^{MOM}$ and $\hat{\theta}^{EL}$ in Example 3.

T	No Outliers			1% from $B(20, \frac{1}{2})$			2% from $N(6, 2)$			2% from $U(6, 7)$		
	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS	EL	MOM	NLS
100	6.700	6.890	6.701	8.559	8.197	8.666	10.502	9.496	10.546	8.876	8.687	8.928
200	3.341	3.351	3.342	5.970	5.446	5.971	8.096	7.911	8.178	7.500	7.438	7.501
300	2.200	2.239	2.201	5.482	5.067	5.483	7.055	6.742	7.088	6.824	6.652	6.825
400	1.565	1.624	1.566	5.102	4.555	5.103	6.673	6.209	6.785	6.528	6.305	6.529
500	1.263	1.295	1.264	4.645	4.188	4.646	6.574	6.045	6.575	6.244	6.111	6.245
600	1.158	1.175	1.159	4.497	3.675	4.498	6.399	5.909	6.400	6.219	6.077	6.220
700	0.831	0.844	0.831	4.248	3.492	4.249	6.290	5.817	6.292	6.101	5.970	6.102
800	0.802	0.815	0.803	4.336	3.486	4.337	6.356	5.664	6.357	6.067	5.691	6.068
900	0.731	0.736	0.732	4.138	3.162	4.139	6.206	5.411	6.207	5.858	5.486	5.859
1000	0.663	0.669	0.664	3.628	2.764	3.629	6.336	5.312	6.336	5.385	4.998	5.386

Form Tables 1–3, we have the following comments.

- (1) The MSE decrease for all estimators as T becomes large whether there are outliers.
- (2) When there are no outliers, the MSE of $\hat{\theta}^{MOM}$, $\hat{\theta}^{NLS}$ and $\hat{\theta}^{EL}$ are the same basically.
- (3) When there are outliers, the MSE of $\hat{\theta}^{MOM}$ estimator is smaller than the MSE of $\hat{\theta}^{NLS}$ estimator and $\hat{\theta}^{EL}$ estimator. From Tables 1 and 3, the results show that they are no significant differences between the MSE of $\hat{\theta}^{NLS}$ estimator and $\hat{\theta}^{EL}$ estimator as T is large.

Example 4. We consider model $y_i = 0.8^{x_i} + \varepsilon_i$, x_i are generated from $P(0.7)$, ε_i are generated from $N(0, 1)$. For the power, we use $\theta + \theta_0$ with $\theta_0 \in \{0.1, 0.2\}$ as the alternative hypothesis. The results are shown in Table 4. MOMEL represents empirical likelihood test based on MOM method, and EL represents hypothesis test based on the EL estimator.

Table 4. Size and power in Example 4.

T	Size		Power			
	MOMEL	EL	$\theta_0 = 0.1$		$\theta_0 = 0.2$	
			MOMEL	EL	MOMEL	EL
100	0.054	0.060	0.629	0.161	0.852	0.584
200	0.051	0.045	0.713	0.237	0.946	0.847
300	0.059	0.062	0.764	0.368	0.987	0.966
400	0.047	0.060	0.828	0.498	0.996	0.987
500	0.050	0.040	0.869	0.578	1.000	0.996
600	0.055	0.047	0.900	0.670	1.000	0.999
700	0.052	0.048	0.920	0.728	1.000	0.999
800	0.050	0.045	0.936	0.779	1.000	1.000
900	0.050	0.046	0.953	0.833	1.000	1.000
1000	0.046	0.040	0.961	0.882	1.000	1.000

Example 5. We consider model $y_i = x_i^{0.6} + \varepsilon_i$, suppose x_i are generated from $U(2, 3)$, ε_i are generated from $N(0, 1)$. For power, we use $\theta + \theta_0$ with $\theta_0 \in \{0.1, 0.15\}$ as the alternative hypothesis. The results are shown in Table 5.

Table 5. Size and power in Example 5.

T	Size		Power			
	MOMEL	EL	$\theta_0 = 0.1$		$\theta_0 = 0.15$	
			MOMEL	EL	MOMEL	EL
100	0.057	0.065	0.762	0.322	0.896	0.524
200	0.050	0.046	0.888	0.556	0.974	0.827
300	0.054	0.043	0.939	0.729	0.995	0.970
400	0.051	0.044	0.972	0.835	0.997	0.991
500	0.055	0.047	0.988	0.919	0.999	0.998
600	0.056	0.047	0.995	0.949	1.000	0.999
700	0.047	0.044	0.998	0.966	1.000	1.000
800	0.056	0.043	1.000	0.985	1.000	1.000
900	0.050	0.048	1.000	0.991	1.000	1.000
1000	0.051	0.046	1.000	0.996	1.000	1.000

Example 6. We consider model $y_i = e^{(0.5x_i)} + \varepsilon_i$, x_i are generated by $U(-1, 0)$, ε_i are generated by $N(0, 1)$. For power, we use $\theta + \theta_0$ with $\theta_0 \in \{0.2, 0.3\}$ as the alternative hypothesis. The results are shown in Table 6.

From simulation results that are displayed in Tables 4–6, we can see that the size of the proposed test is close to 0.05 and the power is close to 1 as T increases. Especially when N is small, the results of MOM are significantly better than EL's. When T increases, the MOM also performs better in terms of size and power although the power of both methods tends to one. In summary, our method is better.

Table 6. Size and power in Example 6.

T	Size		Power			
	MOMEL	EL	$\theta_0 = 0.2$		$\theta_0 = 0.3$	
			MOMEL	EL	MOMEL	EL
100	0.063	0.072	0.577	0.144	0.663	0.190
200	0.058	0.045	0.640	0.183	0.754	0.290
300	0.047	0.040	0.698	0.215	0.834	0.398
400	0.051	0.049	0.768	0.291	0.885	0.501
500	0.055	0.047	0.799	0.335	0.892	0.577
600	0.057	0.042	0.824	0.373	0.946	0.667
700	0.056	0.043	0.851	0.470	0.957	0.725
800	0.048	0.049	0.859	0.472	0.969	0.801
900	0.055	0.042	0.893	0.540	0.983	0.852
1000	0.051	0.057	0.920	0.607	1.000	0.880

5. The Real Data Analysis

In this section, we apply the MOM method to analyze the top 50 data of GDP of China in 2019. Basing on the presentation of Zhu et al. [17], there are many methods to test whether there are outliers in the data, such as the 4d test, 3σ principle, the Chauvenet method, the t-test and the Grubbs test. Sun [18] also introduced the box plot method. Different test methods will get different outliers. So we use the box plot as shown in Figure 1 to confirm the existence of outliers in the actual data based on the suggestion of Sun et al. [18]. The outliers are 381.55, 353.71, 269.27, and 236.28 (unit: ten billion RMB).

We also use a 3σ principle to test whether there are outliers, and the result shows that the outliers are 381.55 and 353.71. Through the test of the above two methods, we can judge that there are outliers in this real data.

Yin and Du [19] introduced a power-law distribution. For the purpose of predicting the GDP development trend of major cities accurately in China. We use the EL method, the MOM method and the NLS method to fit the curve respectively. Where x_i represents the sorting order of GDP of 50 cities in descending order. The dataset is from www.askci.com (accessed on 15 February 2021).

The EL gives the nonlinear regression equation

$$GDP^{EL} = 444.0250 \times x_i^{-0.5176290}. \quad (9)$$

The MOM gives the nonlinear regression equation

$$GDP^{MOM} = 594.1439 \times x_i^{-0.6111023}. \quad (10)$$

The NLS gives the nonlinear regression equation

$$GDP^{NLS} = 443.0247 \times x_i^{-0.5167945}. \quad (11)$$

In Figure 2, the red line represents the fitting result of NLS method, and the blue line represents the fitting result of MOM method. the black line represents the fitting result of EL method, and the yellow points represent the true value of GDP.

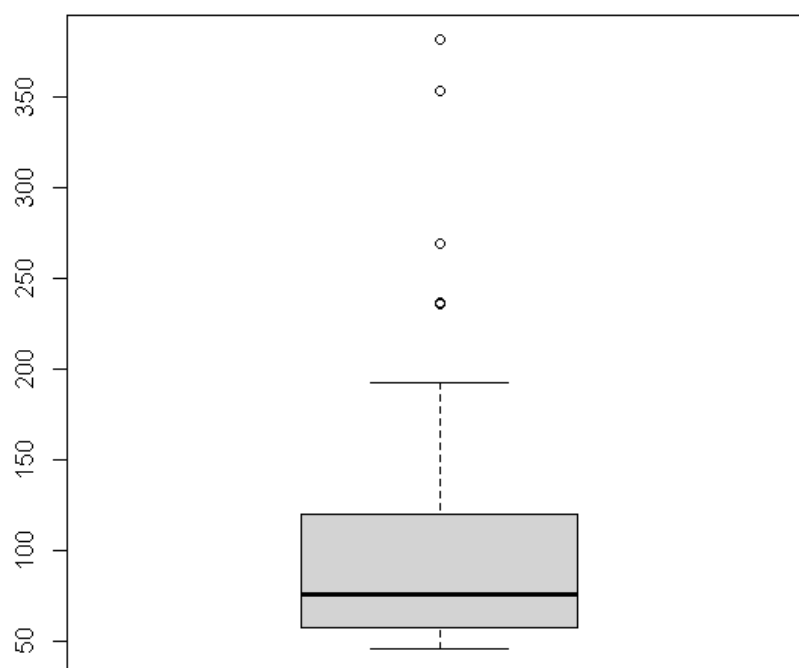


Figure 1. Box plot of the top 50 data of GDP of China in 2019.

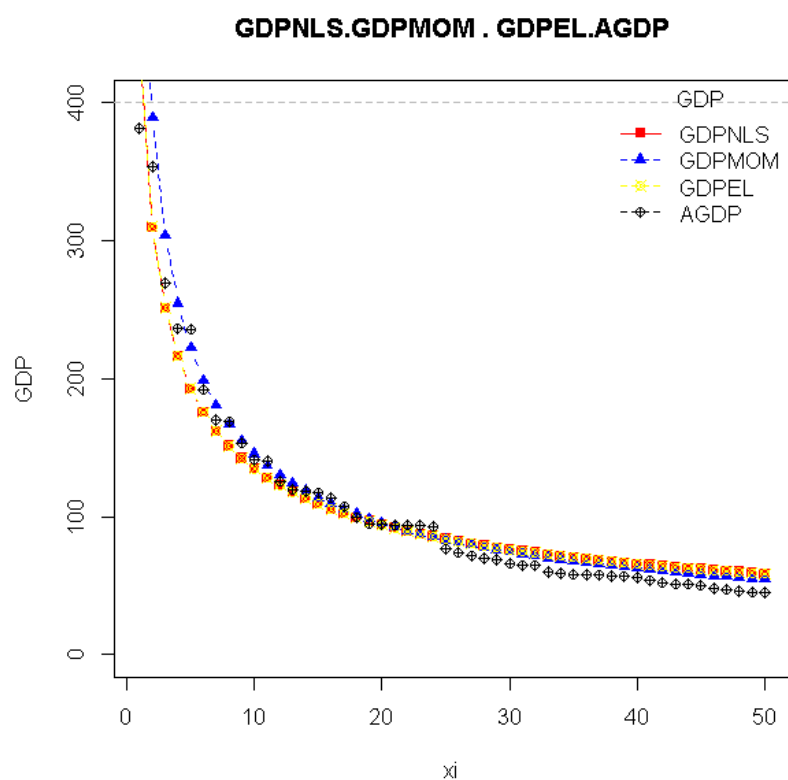


Figure 2. Fitting result figure of the top 50 data of GDP of China in 2019.

In actual data, the true values of parameters are really unknown, so we cannot calculate MSE of the parameter. MAE refers to the average value of the absolute error. The definition of Mean Absolute Error (MAE) is given below.

$$MAE = \frac{1}{G} \sum_{i=1}^G |y_i - \hat{y}_i|, \quad i = 1, \dots, G.$$

In the actual data, y_i refers to the true value of GDP, and \hat{y}_i refers to the GDP value obtained from the fitted nonlinear regression model, so we calculated the MAE. MAE of the MOM method is 11.984. MAE of the NLS method is 12.024, MAE of the MOM method is 11.982. Cross-validations are taken to examine the accuracy of forecasting. Specifically, we take 40 data as experimental data and the other 10 as forecasting data randomly and the number of independent replications is 1000. The MAE of EL, ELS and MOM are 14.206, 14.271 and 12.242 respectively. These suggest that MOM is more plausible than NLS and EL.

6. Conclusions

It is shown that the NLS method is not robust to outliers based on the research of Gao and Li [11]. So in this paper, firstly, we apply the MOM method to the nonlinear regression model and introduce its theory. We give the theoretical results of asymptotic normality and consistency of the MOM estimator. Secondly, we propose a new test method based on the empirical likelihood method. Thirdly, we use the MOM method to estimate the parameters of three forms of nonlinear regression models, and compare the MSE of $\hat{\theta}^{NLS}$, $\hat{\theta}^{MOM}$ and $\hat{\theta}^{EL}$. The results show that the MSE of $\hat{\theta}^{MOM}$ is the smallest from Tables 1–3 and the size and power prove the superiority of the MOM method from Tables 4–6. Finally, the MOM method is applied to predict the GDP development of cities of China, the value of MAE shows that the prediction of the MOM method is better than the NLS method. All in all, the MOM method does not need to eliminate outliers. Regardless of whether there are outliers in the data, we will use the MOM method to get a robust estimation.

Author Contributions: Conceptualization, P.L.; methodology, P.L.; M.Z. and Q.Z.; software, R.Z.; writing—original draft, R.Z.; writing—review and editing, M.Z., R.Z. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Pengfei Liu’s research is supported by the National Natural Science Foundation of China (NSFC11501261, NSFC52034007) and the State Scholarship funded by China Scholarship Council (CSC201808320107). Ru Zhang’s research is supported by the Project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. Qin Zhou’s research is supported by the National Natural Science Foundation of China (NSFC11671178).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: www.askci.com (accessed on 15 February 2021).

Acknowledgments: Thank Shaochen Wang and Wang Zhou for their help. And thank the reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this Appendix A, we give the technical proofs of Theorems 1–3.

Lemma A1 (Chernoff's inequality, cf. Vershynin [20] Theorem 2.3.1). Let $X_i (i = 1, \dots, n)$ is an independent Bernoulli random variable with the parameter p_i . Consider sum $M_n = \sum_{i=1}^n X_i$ and their mean $\mu = \mathbb{E}(M_n)$, for any $t > \mu$, we have

$$\mathbb{P}(M_n \geq t) \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t. \quad (\text{A1})$$

Proof of Theorem 1. In accordance with the condition (I) of Theorem 1 and the Lemma 1 of Ivanov [21], for $n \geq N_0$ and sufficiently large positive ρ , c does not depend on n and ρ . We have

$$\mathbb{P}(\sqrt{n}|\hat{\theta}_k^{(j)} - \theta_k| > \rho) \leq \frac{c}{\rho^s}, \quad j = 1, \dots, g. \quad (\text{A2})$$

According to Wu [22], we can get that the least square estimate of σ^2 is ($j = 1, \dots, g$)

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\theta}_q^{(j)})^2, \quad q = 1, \dots, k$$

According to Formula (2) and the conditions (II)(III)(IV). From the Theorem 5 of Wu [12], we can know

$$\sqrt{n}(\hat{\theta}_q^{(j)} - \theta_q) \xrightarrow{d} N(0, \hat{\sigma}_n^2 S^{-1}), \quad q = 1, \dots, k. \quad (\text{A3})$$

According to Pinelis [23], C_1 is a constant, we can know

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\sqrt{n}}{\hat{\sigma}_n S^{\frac{-1}{2}}} (\hat{\theta}_q^{(j)} - \theta_q) \leq x \right) - \Phi(x) \right| \leq \frac{C_1}{\sqrt{n}}, \quad q = 1, \dots, k. \quad (\text{A4})$$

where Φ represent the cumulative distribution function of Standard normal distribution. Define random variables

$$\alpha_{n,j} = \frac{\sqrt{n}}{\hat{\sigma}_n S^{\frac{-1}{2}}} (\hat{\theta}_1^{(j)} - \theta_1), \quad j = 1, \dots, g. \quad (\text{A5})$$

According to formula (A4), we have

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\sqrt{n}}{\hat{\sigma}_n S^{\frac{-1}{2}}} (\hat{\theta}_1^{(j)} - \theta_1) \leq x \right) - \Phi(x) \right| \leq \frac{C_1}{\sqrt{n}}$$

So we can get

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\alpha_{n,j} \leq x) - \Phi(x)| \leq \frac{C_1}{\sqrt{n}}$$

For each $j = 1, \dots, g$, suppose $x = \frac{\sqrt{n}H}{\hat{\sigma}_n S^{-\frac{1}{2}}}$, we have

$$\mathbb{P}(\hat{\theta}_1^{(j)} - \theta_1 \geq H) \leq \frac{C_1}{\sqrt{n}} + 1 - \Phi\left(\frac{\sqrt{n}H}{\hat{\sigma}_n S^{-\frac{1}{2}}}\right)$$

for all $H > 0$, according to the elementary inequality

$$1 - \Phi(\sqrt{n}H/\hat{\sigma}_n S^{-\frac{1}{2}}) \leq e^{-nH^2/2(\hat{\sigma}_n^2 S^{-1})}$$

where $o(n^{-\frac{1}{2}})$ for large n and fixed $H > 0$, hence

$$\mathbb{P}(\hat{\theta}_1^{(j)} - \theta_1 \geq H) \leq \frac{C_2}{2\sqrt{n}}. \quad (\text{A6})$$

Similarly, we can get

$$\mathbb{P}(\hat{\theta}_1^{(j)} - \theta_1 \leq -H) \leq \frac{C_2}{2\sqrt{n}}$$

where C_2 is a constant that depends on H but not n , so we have

$$\mathbb{P}(|\hat{\theta}_1^{(j)} - \theta_1| \geq H) \leq \frac{C_2}{\sqrt{n}}. \quad (\text{A7})$$

It is easy to verify that

$$|\hat{\theta}_1^{MOM} - \theta_1| \leq \text{Median}\{|\hat{\theta}_1^{(j)} - \theta_1|, j = 1, \dots, g\}$$

so we have the conclusion

$$\mathbb{P}(|\hat{\theta}_1^{MOM} - \theta_1| \geq H) \leq \mathbb{P}(\text{Median}\{|\hat{\theta}_1^{(j)} - \theta_1|, j = 1, \dots, g\} \geq H) := \mathbb{P}(E). \quad (\text{A8})$$

Define Bernoulli random variable

$$\eta_j = I(|\hat{\theta}_1^{(j)} - \theta_1| \geq H), \quad j = 1, \dots, g$$

we have $E\eta_j \leq C_2 n^{-\frac{1}{2}}$ by Formula (A9). It can be seen that event E occurs if and only if $\sum_{j=1}^g \eta_j$ is larger than $\frac{g}{2}$, hence

$$\mathbb{P}(E) = \mathbb{P}\left(\sum_{j=1}^g \eta_j \geq \frac{g}{2}\right) \leq e^{-gE\eta_1} \left(2eC_2 n^{-\frac{1}{2}}\right)^{\frac{g}{2}} \leq \frac{C}{n^{\frac{g}{5}}}. \quad (\text{A9})$$

We have used Lemma 1 in the last step. This ends the proof of Theorem 1.

For any fixed x , we define i.i.d random variables

$$\pi_{n,j}(x) = I(\alpha_{n,j} \leq x), \quad j = 1, \dots, g$$

and suppose

$$p_n(x) = \mathbb{P}(\alpha_{n,j} \leq x)$$

according to Formula (A4)

$$|p_n(x) - \Phi(x)| = O(n^{-\frac{1}{2}})$$

for all real x . The following lemma gives the central limit theorem for the partial sums of $\pi_{n,j}(x)$. \square

Lemma A2. Suppose $n/g \rightarrow \infty$ as $g \rightarrow \infty$. We have

$$\sqrt{g} \left(\frac{1}{g} \sum_{j=1}^g \pi_{n,j}(x) - \Phi(x) \right) \xrightarrow{d} N(0, \Phi(x)[1 - \Phi(x)]). \quad (\text{A10})$$

for the fixed x , as $g \rightarrow \infty$,

$$\sqrt{g} \left(\frac{1}{g} \sum_{j=1}^g \pi_{n,j}(xg^{-\frac{1}{2}}) - \frac{1}{2} - \frac{x}{\sqrt{2\pi g}} \right) \xrightarrow{d} N(0, \frac{1}{4}). \quad (\text{A11})$$

Proof of Lemma 2. For convenience, we write $\pi_{n,j}(x)$ as $\pi_{n,j}$. By independence, for any real t and $i = \sqrt{-1}$, we have

$$\mathbb{E} \exp \left\{ it \sqrt{g} \left(\frac{1}{g} \sum_{j=1}^g \pi_{n,j} - \Phi(x) \right) \right\} = \left(\mathbb{E} e^{it \frac{1}{\sqrt{g}} [\pi_{n,1} - \Phi(x)]} \right)^g. \quad (\text{A12})$$

through the Taylor's expansion, we have

$$\begin{aligned} \mathbb{E} e^{it \frac{1}{\sqrt{g}} [\pi_{n,1} - \Phi(x)]} &= p_n e^{it g^{-\frac{1}{2}} (1 - \Phi(x))} + (1 - p_n) e^{-it g^{-1/2} \Phi(x)} \\ &= p_n (1 + it g^{-1/2} (1 - \Phi(x)) + \frac{(it g^{-1/2} (1 - \Phi(x)))^2}{2!}) \\ &\quad + (1 - p_n) (1 - it g^{-1/2} \Phi(x) + \frac{(-it g^{-1/2} \Phi(x))^2}{2!}) \\ &= 1 + p_n it g^{-1/2} (1 - \Phi(x)) - (1 - p_n) it g^{-1/2} \Phi(x) \\ &\quad - \frac{p_n}{2g} [t(1 - \Phi(x))]^2 - \frac{1 - p_n}{2g} [t\Phi(x)]^2 + o(g^{-1}) \\ &= 1 - \frac{t^2}{2g} \Phi(x)[1 - \Phi(x)] + o(g^{-1}). \end{aligned} \quad (\text{A13})$$

where we used the formula $|p_n - \Phi(x)| = O(n^{-\frac{1}{2}})$, when $n/g \rightarrow \infty$ and $g \rightarrow \infty$,

$$|p_n g^{-1/2} (1 - \Phi(x)) - (1 - p_n) g^{-1/2} \Phi(x)| = g^{-1/2} |p_n - \Phi(x)| = o(g^{-1}). \quad (\text{A14})$$

so the first conclusion of the Lemma 2 can get by formula (A13).

For the second conclusion, we find that the above calculations still hold if we replace x with $xg^{-\frac{1}{2}}$ and note the fact that

$$\Phi(xg^{-\frac{1}{2}}) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{xg^{-1/2}} e^{-\frac{u^2}{2}} du = \frac{1}{2} + \frac{x}{\sqrt{2\pi g}} + o(g^{-\frac{1}{2}}). \quad (\text{A15})$$

We can proof the formula (A15) by the virtue of Slutsky's theorem. \square

Proof of Theorem 2. (1) This follows immediately by formula (A4) and the continuous mapping theorem since the Median function is continuous.

(2) We can observe that

$$\frac{\sqrt{N}}{\hat{\sigma}_n S^{-\frac{1}{2}}} (\hat{\theta}_1^{\text{MOM}} - \theta_1) = \sqrt{g} \frac{\sqrt{n}}{\hat{\sigma}_n S^{-\frac{1}{2}}} (\hat{\theta}_1^{\text{MOM}} - \theta_1) = \sqrt{g} \text{Median}\{\alpha_{n,j}, j = 1, \dots, g\}$$

We first assume g is odd and for any real x , and we have

$$\begin{aligned} &\mathbb{P}(\sqrt{g} \text{Median}\{\alpha_{n,j}, j = 1, \dots, g\} \leq x) \\ &= \mathbb{P}\left(\sum_{j=1}^g I(\alpha_{n,j} \leq xg^{-\frac{1}{2}}) \geq (g+1)/2\right) \\ &= \mathbb{P}\left(\sqrt{g} \left\{ \frac{1}{g} \sum_{j=1}^g \pi_{n,j}(xg^{-\frac{1}{2}}) - 1/2 - \frac{x}{\sqrt{2\pi g}} \right\} \geq -\frac{x}{2\sqrt{\pi}} + O(g^{-\frac{1}{2}})\right). \end{aligned} \quad (\text{A16})$$

under the above lemma, it tends to $\Phi(\sqrt{\frac{2}{\pi}}x)$.

If g is even, we can know

$$\mathbb{P}(\sqrt{g} \text{Median}\{\alpha_{n,j}, j = 1, \dots, g\} \leq x) \geq \mathbb{P}\left(\sum_{j=1}^g I(\alpha_{n,j} \leq xg^{-\frac{1}{2}}) \geq \frac{g}{2} + 1\right)$$

and

$$\mathbb{P}(\sqrt{g} \text{Median}\{\alpha_{n,j}, j = 1, \dots, g\} \leq x) \leq \mathbb{P}\left(\sum_{j=1}^g I(\alpha_{n,j} \leq xg^{-\frac{1}{2}}) \geq \frac{g}{2}\right)$$

The right hand sides of the above two inequalities tend to $\Phi(\sqrt{\frac{2}{\pi}}x)$ as $g \rightarrow \infty$. \square

Proof of Theorem 3. Recall that

$$T_{n,j} = I(\hat{\theta}_q^{(j)} \leq \theta_q), \quad q = 1, \dots, k. \quad (\text{A17})$$

where $j = 1, \dots, g$, so formula (6) is

$$f(\lambda) = \frac{1}{g} \sum_{j=1}^g \frac{T_{n,j} - 0.5}{1 + \lambda(T_{n,j} - 0.5)} = 0. \quad (\text{A18})$$

set that $L_{n,j} = \lambda(T_{n,j} - 0.5)$, and we have

$$\begin{aligned} \lambda \tilde{R} &= \frac{1}{g} \sum_{j=1}^g \frac{\lambda(T_{n,j} - 0.5)^2}{1 + L_{n,j}} \\ &= \frac{1}{g} \sum_{j=1}^g \frac{L_{n,j}(T_{n,j} - 0.5)}{1 + L_{n,j}}. \end{aligned} \quad (\text{A19})$$

$$\begin{aligned} \bar{T}_{n,j} - 0.5 &= \frac{1}{g} \sum_{j=1}^g \frac{1 + L_{n,j}}{1 + L_{n,j}} (T_{n,j} - 0.5) \\ &= \frac{1}{g} \sum_{j=1}^g (T_{n,j} - 0.5) - \frac{1}{g} \sum_{j=1}^g \frac{T_{n,j} - 0.5}{1 + L_{n,j}} \\ &= \frac{1}{g} \sum_{j=1}^g \frac{L_{n,j}(T_{n,j} - 0.5)}{1 + L_{n,j}}. \end{aligned} \quad (\text{A20})$$

So

$$(\bar{T}_{n,j} - 0.5) = \lambda \tilde{R}. \quad (\text{A21})$$

where

$$\begin{aligned} \tilde{R} &= \frac{1}{g} \sum_{j=1}^g \frac{(T_{n,j} - 0.5)^2}{1 + L_{n,j}} \\ \bar{T}_{n,j} &= \frac{1}{g} \sum_{j=1}^g T_{n,j} \\ R &= \frac{1}{g} \sum_{j=1}^g (T_{n,j} - 0.5)^2 = 0.25. \\ T_g &= \max_{1 \leq j \leq g} |T_{n,j} - 0.5| = 0.5 \end{aligned} \quad (\text{A22})$$

Combining the constraint condition $\omega_i > 0$, we can get that $1 + L_{n,j} > 0$, and

$$\begin{aligned}\lambda R &\leq \lambda \tilde{R}(1 + \max_{1 \leq j \leq g} L_{n,j}) \\ &\leq \lambda \tilde{R}(1 + \lambda T_g) \\ &= (\bar{T}_{n,j} - 0.5)(1 + \lambda T_g)\end{aligned}$$

The last equality follows by formal (A20). So,

$$\lambda[R - (\bar{T}_{n,j} - 0.5)T_g] \leq \bar{T}_{n,j} - 0.5. \quad (\text{A23})$$

and according to Lemma 2, $\bar{T}_{n,j} - 0.5 = O_p(g^{-1/2})$, we can get

$$\lambda[0.25 - O_p(g^{-1/2})] = O_p(g^{-1/2})$$

so

$$\lambda = O_p(g^{-1/2}). \quad (\text{A24})$$

In addition, we know

$$\max_{1 \leq j \leq g} |L_{n,j}| = O_p(g^{-1/2}). \quad (\text{A25})$$

Expanding formula (6),

$$\begin{aligned}0 &= \frac{1}{g} \sum_{j=1}^g \frac{T_{n,j} - 0.5}{1 + L_{n,j}} \\ &= (\bar{T}_{n,j} - 0.5) - \lambda R + \frac{1}{g} \sum_{j=1}^g \frac{(T_{n,j} - 0.5)L_{n,j}^2}{1 + L_{n,j}} \\ &= (\bar{T}_{n,j} - 0.5) - 0.25\lambda + \frac{1}{g} \sum_{j=1}^g \frac{(T_{n,j} - 0.5)L_{n,j}^2}{1 + L_{n,j}}.\end{aligned} \quad (\text{A26})$$

The final term in formula (A26) above has a norm bounded by

$$\frac{1}{g} \sum_{j=1}^g |T_{n,j} - 0.5|^3 \lambda^2 |1 + L_{n,j}|^{-1} = O(1)(O_p(g^{-1/2}))^2 O_p(1) = o_p(g^{-1/2})$$

Therefore

$$\lambda = R^{-1}(\bar{T}_{n,j} - 0.5) + \beta = 4(\bar{T}_{n,j} - 0.5) + \beta$$

where $\beta = o_p(g^{-1/2})$.

Through formula (A26) and using Taylor expansion, we can find that

$$\log(1 + L_{n,j}) = L_{n,j} - \frac{1}{2}L_{n,j}^2 + \eta_j. \quad (\text{A27})$$

holds for some finite $B > 0$, $1 \leq j \leq g$,

$$\mathbb{P}(|\eta_j| \leq B|L_{n,j}|^3) \rightarrow 1. \quad (\text{A28})$$

as $g \rightarrow \infty$ and $n \rightarrow \infty$.

Now, we calculate that

$$\begin{aligned}
 -2\log R(\theta) &= 2 \sum_{j=1}^g \log(1 + L_{n,j}) \\
 &= 2 \sum_{j=1}^g (L_{n,j} - \frac{1}{2} L_{n,j}^2 + \eta_j) \\
 &= \sum_{j=1}^g (2(4(\bar{T}_{n,j} - 0.5) + \beta) T_{n,j} \\
 &\quad - (4\bar{T}_{n,j} - 2 + \beta) - (4\bar{T}_{n,j} - 2 + \beta)^2 T_{n,j}^2 \\
 &\quad + (4\bar{T}_{n,j} - 2 + \beta)^2 T_{n,j} - 0.25(4\bar{T}_{n,j} - 2 + \beta)^2 + 2 \sum_{j=1}^g \eta_j) \\
 &= 4g(\bar{T}_{n,j} - 0.5)^2 - \frac{1}{4}g\beta^2 + 2 \sum_{j=1}^g \eta_j
 \end{aligned}$$

By Lemma 2, we have

$$4g(\bar{T}_{n,j} - 0.5)^2 \rightarrow \chi_1^2. \quad (\text{A29})$$

Noticed that

$$\frac{1}{4}g\beta^2 = \frac{1}{4}g o_p(g^{-1}) = o_p(1). \quad (\text{A30})$$

$$|\sum_{j=1}^g \eta_j| \leq B|\lambda|^3 \sum_{j=1}^g |T_{n,j} - 0.5|^3 = O_p(g^{-\frac{3}{2}})O(1) = o_p(1)$$

This completes the proof. \square

References

- Hong, Z. The application of nonlinear regression model to the economic system prediction. *J. Jimei Inst. Navig.* **1996**, *4*, 48–52.
- Wang, D.; Jiang, D.; Cheng, S. Application of nonlinear regression model to detect the thickness of protein layer. *J. Biophys.* **2000**, *16*, 33–74.
- Chen, H.; Wang, J.; Zhang, H. Application of nonlinear regression analysis in establishing price model of ground-to-air missile. *J. Abbr.* **2005**, *4*, 77–79.
- Archontoulis, S.V.; Miguez, F.E. Nonlinear regression models and applications in agricultural research. *Agron. J.* **2015**, *105*, 1–13. [\[CrossRef\]](#)
- Alon, N.; Matias, Y.; Szegedy, M. The space complexity of approximating the frequency moment. *J. Comput. Syst. Sci.* **1999**, *58*, 137–147. [\[CrossRef\]](#)
- Lecué, G.; Lerasle, M. Robust machine learning by median-of-means: Theory and practice. *Ann. Stat.* **2017**, *32*, 4711–4759.
- Lecué, G.; Lerasle, M.; Mathieu, T. Robust classification via MOM minimization. *Mach. Learn.* **2018**, *32*, 1808–1837. [\[CrossRef\]](#)
- Zhang, Y.; Liu, P. Median-of-means approach for repeated measures data. *Commun. Stat. Theory Methods* **2020**, *2020*, 1–10. [\[CrossRef\]](#)
- Radchenko, P.P. Nonlinear least-squares estimation. *J. Multivar. Anal.* **2006**, *97*, 548–562.
- Ding, X.; Xu, L.; Lin, J. Empirical likelihood diagnosis of nonlinear regression model. *Chin. J. Appl. Math.* **2012**, *4*, 693–702.
- Gao, S.; Li, X. Analysis on the robustness of least squares method. *Stat. Decis.* **2006**, *15*, 125–126.
- Wu, C.-F. Asymptotic theory of nonlinear least squares estimation. *Ann. Stat.* **1981**, *9*, 501–513. [\[CrossRef\]](#)
- Emilien, J.; Gábor, L.; Roberto, I.O. Sub-Gaussian estimators of the mean of a random vector. *Ann. Stat.* **2017**, *47*, 440–451.
- Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249. [\[CrossRef\]](#)
- Jiang, Y. Empirical Likelihood Inference of Nonlinear Regression Model Parameters. Master's Thesis, Beijing University of Technology, Beijing, China, 2005.
- Ratkowski, H.Z. *Nonlinear Regression Model: A Unified Practical Method*; Nanjing University Press: Nanjing, China, 1986; pp. 12–25.
- Zhu, J.; Bao, Y.; Li, C. Discussion on data outlier test and processing method. *Univ. Chem.* **2018**, *33*, 58–65. [\[CrossRef\]](#)

18. Sun, X.; Liu, Y.; Chen, W.; Jia, Z.; Huang, B. The application of box and plot method in the outlier inspection of animal health data. *China Anim. Quar.* **2010**, *27*, 66–68.
19. Yin, C.; Du, J. The collision theory reaction rate coefficient for power-law distributions. *Phys. A Stat. Mech. Its Appl.* **2014**, *407*, 119–127. [[CrossRef](#)]
20. Vershynin, R. *High-Dimensional Probability (An Introduction with Applications in Data Science)*; Cambridge University Press: Cambridge, UK, 2018; pp. 70–97.
21. Ivanov, A.V. An asymptotic expansion for the Distribution of the Least Squares Estimator of the nonlinear regression parameter. *Theory Probab. Appl.* **1977**, *21*, 557–570. [[CrossRef](#)]
22. Wu, Q. Asymptotic normality of least squares estimation in nonlinear models. *J. Guilin Inst. Technol.* **1998**, *18*, 394–400.
23. Pinelis, I.; Molzon, R. Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electron. J. Stat.* **2016**, *10*, 1001–1063. [[CrossRef](#)]