Filomat 38:7 (2024), 2521–2535
https://doi.org/10.2298/FIL2407521F

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Serial correlation test of parametric regression models with response missing at random

Guoliang Fan[a], Jie Ji[a], Hongxia Xu[b,*]

[a]School of Economics and Management, Shanghai Maritime University, Shanghai, 201306, China
[b]School of Science, Shanghai Maritime University, Shanghai, 201306, China

Abstract. It is well-known that successive residuals may be correlated with each other, and serial correlation usually result in an inefficient estimate in time series analysis. In this paper, we investigate the serial correlation test of parametric regression models where the response is missing at random. Three test statistics based on the empirical likelihood method are proposed to test serial correlation. It is proved that three proposed empirical likelihood ratios admit limiting chi-square distribution under the null hypothesis of no serial correlation. The proposed test statistics are simple to calculate and convenient to use, and they can test not only zero first-order serial correlation, but also the higher-order serial correlation. A simulation study and a real data analysis are conducted to evaluate the finite sample performance of our proposed test methods.

1. Introduction

Testing for serial correlation has long been a standard practice in applied econometric analysis because if the residuals are serially correlated, not only the least squares estimator is inefficient, it can be inconsistent if the regressors contain lagged dependent variables. Hence, it is important to test the serial correlation of the residuals before using the models. There exist some approaches to test serial correlation. For example, Chi and Reinsel ([6], [7]) proposed score test based on likelihood function. This kind of test method usually has good power when the error follows the normal distribution; see [37] and [12] for more details about score test. However, when the normal distribution assumption is violated, the effect of the test is unsatisfactory. In addition, the likelihood function is quite complex for the case of high order serial correlation, so the test method is not easy to be extended to high order serial correlation test case. Another method is based on the least squares residuals test, such as the well-known Durbin-Watson test, which is only suitable for testing first-order serial correlation. Besides, [25] studied tests for serial correlation in semiparametric partially linear panel data model and [5] extended their results to the pure time series case under certain mixing conditions by the generalization of the Durbin h-statistic. Empirical likelihood (EL) based testing method is an attractive approach for testing serial correlation, see [21] for example.

2020 *Mathematics Subject Classification*. Primary 62D10; Secondary 62G10.
Keywords. Empirical likelihood, Imputation, Inverse probability weighting, Missing at random, Serial correlation test.
Received: 23 December 2022; Accepted: 01 October 2023
Communicated by Aleksandar Nastić
Research supported by the National Social Science Fund of China (22BTJ018).
* Corresponding author: Hongxia Xu
Email addresses: glfan@shmtu.edu.cn (Guoliang Fan), jijie202002@163.com (Jie Ji), xuhx@shmtu.edu.cn (Hongxia Xu)

Parametric regression models are frequently used to describe the association between a response variable and its predictors, and have the following form:

$$Y = f(\mathbf{x}, \theta) + \varepsilon, \tag{1}$$

where $\mathbf{x} = (X_1, \cdots, X_m)^{\mathrm{T}}$ is a vector of the predictors, $f(\cdot, \theta)$ is a known parameter function with unknown parameter vector θ and ε is the model error. Examples of $f(\cdot, \theta)$ include the widely used linear model $f(\mathbf{x}, \theta) = \mathbf{x}^{\mathrm{T}}\theta$ and a logistic model $f(\mathbf{x}, \theta) = 1/\{1 + \exp(\mathbf{x}^{\mathrm{T}}\theta)\}$. Parametric regression models (1) have been extensively studied by many authors. For example, Fan and Huang([15]) investigated the goodness-of-fit test for parametric regression models by using the adaptive Neyman test. Van Keilegom et al.([13]) proposed a test statistic which measures the distance between the empirical distribution function of the parametric and of the nonparametric residuals to study the goodness-of-fit tests in parametric regression models. Dai and Müller([9]) applied nonparametric method to give efficient estimators for expectations of a known function of response and covariates in parametric regression models with responses missing at random (MAR).

The aforementioned articles mainly discussed the serial correlation test when the data are completely observed. In practice, however, not all response measurements are observable due to various reasons such as unwillingness of some sampled units to supply the desired information, loss of information caused by uncontrollable factors, failure on the part of investigators to gather correct information, and so forth. Thus, it is of interest for us to investigate serial correlation test with missing response. The simplest way of dealing with missing response data is to just omit those participants who have any missing data among its variables. Such an analysis is called a complete case (CC) analysis and is proved to be undesirable. It is because the CC analysis may decrease the power of the analysis by decreasing the effective sample size. At present, many authors have studied the processing methods for dealing with missing response data, such as [11] for dimension reduction with MAR and [39] for robust model selection with MAR, among others. In practice, one often obtains a random sample of incomplete data

$$(Y_i, \mathbf{x}_i, \delta_i), \quad i = 1, \cdots, n,$$

where $\delta_i = 0$ if Y_i is missing and $\delta_i = 1$ otherwise. The missing at random (MAR) assumption implies that δ and Y are conditionally independent given $\mathbf{x}$, or equivalently,

$$P(\delta = 1 \mid Y, \mathbf{x}) = P(\delta = 1 \mid \mathbf{x}) \stackrel{\text{def}}{=} \pi(\mathbf{x}),$$

where $\pi(\mathbf{x})$ is called a selection probability function. MAR is a common missing mechanism, which is more general than missing completely at random (MCAR), see [27].

In this paper, our aim is to test the serial correlation for parametric regression model (1), when the response Y is MAR and the covariate $\mathbf{x}$ is completely observed. In order to be able to test the serial correlation both for zero first-order and higher-order cases, we propose three different EL-based test methods to check the possible serial correlation in the model error. The EL method, firstly proposed by Owen (1988), defines an empirical likelihood ratio function, and uses its maximum subject to a hypothesis that place restrictions on the parameter to construct confidence region. This method uses only the data to determine the shape and orientation of a confidence region and does not use the estimator of the asymptotic covariance. The EL method has been used in various test problems. For example, Hu et al.([35]) proposed an empirical log-likelihood ratio to test finite-order serial correlation in semiparametric varying-coefficient partially linear models; Zhou et al.([42]) considered the spline empirical log-likelihood ratio for testing serial correlation in partially nonlinear models; Liu et al.([34]) tested the serial correlation of partial linear models via empirical likelihood ratio; Chen et al.([3]) constructed balanced adjusted empirical likelihood and obtained the asymptotic normality of the empirical log-likelihood ratio statistic when the sample size and the data dimension are comparable; Qin and Zhou([36]) proposed one-sided empirical likelihood method for the complete independence test based on squared sample correlation coefficients.

The rest of this article is organized as follows. In Section 2, we construct three different test statistics based on the EL method. The limiting distributions of the proposed test statistics under the null hypothesis are established in the section too. The performance of the tests via simulation study and a real data analysis are demonstrated in Section 3. The concluding remarks are given in Section 4. All proofs of the main results are given in the Appendix.

2. Methodology

In this section, our interest is to test whether the error ε_i are serially correlated. The null hypothesis to be tested is that the errors ε_i are serially uncorrelated. The alternative hypothesis of interest is a p-th order autoregression, denoted by AR(p) and written as

$$\varepsilon_i = a_1\varepsilon_{i-1} + a_2\varepsilon_{i-2} + \cdots + a_p\varepsilon_{i-p} + e_i,$$

or a p-th order moving average, denoted by MA(p) and written as

$$\varepsilon_i = a_1 e_{i-1} + a_2 e_{i-2} + \cdots + a_p e_{i-p} + e_i,$$

where e_i are i.i.d. with $E(e_i) = 0$ and $\mathrm{var}(e_i) = \sigma^2 < \infty$. Assume that a_i in AR(p) satisfies the stationary condition that the roots of equation $1 - a_1 u - \cdots - a_p u^p = 0$ lie outside the unit circle. Let $\mathbf{a} = (a_1, \cdots, a_p)^{\mathrm{T}}$, then our aim is to test whether $\mathbf{a} = \mathbf{0}_{p\times 1}$ or not. Denote $\boldsymbol{\gamma} = \left(\gamma_1, \gamma_2, \cdots, \gamma_p\right)^{\mathrm{T}}, \gamma_k = E(\varepsilon_i\varepsilon_{i+k})$, $k = 1, 2, \ldots, p$, $i = 1, 2, \ldots, T$, $T = n - p$. By the Yule-Walker equation, it is easy to find that testing whether $\mathbf{a} = \mathbf{0}_{p\times 1}$ or not is equivalent to testing whether $\boldsymbol{\gamma} = \mathbf{0}_{p\times 1}$ or not. Therefore, our hypothesis testing problem becomes to

$$H_0 : \boldsymbol{\gamma} = \mathbf{0}_{p\times 1} \quad \text{vs} \quad H_1 : \boldsymbol{\gamma} \neq \mathbf{0}_{p\times 1}. \tag{2}$$

Let $\mathbf{z}_i = \left(z_{i1}, z_{i2}, \cdots, z_{ip}\right)^{\mathrm{T}}$, where $z_{ik} = \varepsilon_i\varepsilon_{i+k}$, $k = 1, \cdots, p$, $i = 1, 2, \ldots, T$, $T = n - p$. Note that $E(\mathbf{z}_i) = \mathbf{0}_{p\times 1}$ under the null hypothesis H_0. Then testing the serial correlation is equivalent to testing whether $E(\mathbf{z}_i)$ is equal to 0. However $z_{ik} = \{Y_i - f(\mathbf{x}_i, \theta)\}\{Y_{i+k} - f(\mathbf{x}_{i+k}, \theta)\}$ cannot be used directly, since the responses Y_i are MAR and z_{ik} contain unknown parameter vector θ.

To deal with the MAR, we need to estimate the selection probability function $\pi(\mathbf{x})$ beforehand. The Nadaraya-Watson estimation approach is often used to estimate $\pi(\mathbf{x})$. However, a fully nonparametric estimation may suffer from the curse of dimensionality and hence unattractive, since the estimation precision decreases rapidly as the dimension of $\mathbf{x}$ increases. In this situation, a parametric approach might be more suitable to estimate $\pi(\mathbf{x})$. Suppose that $\pi(\mathbf{x})$ has a parametric structure $\pi(\mathbf{x}) = \pi(\mathbf{x}, \boldsymbol{\alpha})$, then one only need to estimate the unknown parameter $\boldsymbol{\alpha}$. Specifically, we assume $\pi(\mathbf{x}, \boldsymbol{\alpha})$ has the following logistic regression,

$$\pi(\mathbf{x}_i, \boldsymbol{\alpha}) = \frac{1}{1 + \exp\left(-\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i\right)}, \quad i = 1, \cdots, n,$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m)^{\mathrm{T}}$ is unknown parameter vector. By the maximum likelihood estimation method, we can obtain the consistent maximum likelihood estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. Then, the estimator of the selection probability function is given as $\widehat{\pi}(\mathbf{x}) = \pi\left(\mathbf{x}, \widehat{\boldsymbol{\alpha}}\right) = \{1 + \exp(-\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{x})\}^{-1}$.

Below we employ the least-squares together with the inverse probability weight (IPW) method to obtain the estimator of the unknown parameter θ, which is given by

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)} \left\{Y_i - f\left(\mathbf{x}_i, \theta\right)\right\}^2.$$

Then, we can construct EL ratio function based on IPW method as follows,

$$\widetilde{R}^{(1)} = \max\left\{\prod_{i=1}^{T} T p_i \;\middle|\; \sum_{i=1}^{T} p_i \widetilde{\mathbf{z}}_i^{(1)} = 0, \sum_{i=1}^{T} p_i = 1, p_i \geq 0\right\}, \tag{3}$$

where $\widetilde{\mathbf{z}}_i^{(1)} = \left(\widetilde{z}_{i1}^{(1)}, \widetilde{z}_{i2}^{(1)}, \cdots, \widetilde{z}_{ip}^{(1)}\right)^{\mathrm{T}}$ and $\widetilde{z}_{ik}^{(1)} = \widehat{\varepsilon_i \varepsilon}_{i+k} = \frac{\delta_i \delta_{i+k}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})}\{Y_i - f(\mathbf{x}_i, \widehat{\theta})\}\{Y_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\}$ for $k = 1, 2, \ldots, p$, $i = 1, 2, \ldots, T$. By the Lagrange multiplier method, the optimal solution of p_i is given by $p_i = T^{-1}\left\{1 + \lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)}\right\}^{-1}$, $\quad i = 1, 2, \cdots, T$, where λ is the solution of the following equation,

$$\sum_{i=1}^{T} \frac{\widetilde{\mathbf{z}}_i^{(1)}}{1 + \lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)}} = 0. \tag{4}$$

The corresponding empirical log-likelihood ratio is then

$$-2\log \widetilde{R}^{(1)} = 2\sum_{i=1}^{T} \log\{1 + \lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)}\}. \tag{5}$$

Theorem 2.1. *Suppose that conditions (C1)-(C4) in Appendix hold. Then under the null hypothesis H_0, we have*

$$-2\log \widetilde{R}^{(1)} \xrightarrow{d} \chi_p^2, \ \ as \ \ n \to \infty,$$

where $\xrightarrow{d}$ stands for convergence in distribution and χ_p^2 is the central chi-square distribution with p degrees of freedom.

Although IPW method is popular for solving the problem of response is MAR, it may not work well when the selection probability is very small or the dimension of covariates is very high. In this circumstance, the samples have a significant impact on the weighted averages due to the outrageous weights, which will result in the biased estimation of sampling distribution. Therefore, this paper will also construct two other test statistics based on imputation method for comparison. During the past research history, the imputation is extraordinarily prevalent approach in dealing with the missing response, see [22] and [18] for example. The directed purpose of imputation is to substitute the missing data with estimated values. Let $\widetilde{Y}_i = \delta_i Y_i + (1 - \delta_i) f(\mathbf{x}_i, \widehat{\theta})$ be the imputation estimator of Y_i. Then we can construct the second EL ratio function based on imputation method as follows,

$$\widetilde{R}^{(2)} = \max\left\{\prod_{i=1}^{T} Tp_i \,\Big|\, \sum_{i=1}^{T} p_i \widetilde{\mathbf{z}}_i^{(2)} = 0, \sum_{i=1}^{T} p_i = 1, p_i \geq 0\right\}, \tag{6}$$

where $\widetilde{\mathbf{z}}_i^{(2)} = \left(\widetilde{z}_{i1}^{(2)}, \widetilde{z}_{i2}^{(2)}, \cdots, \widetilde{z}_{ip}^{(2)}\right)^{\mathrm{T}}$ and $\widetilde{z}_{ip}^{(2)} = \{\widetilde{Y}_i - f(x_i, \widehat{\theta})\} \cdot \{\widetilde{Y}_{i+k} - f(x_{i+k}, \widehat{\theta})\}$, $i = 1, \cdots, T$, $k = 1, \cdots, p$. Then the empirical log-likelihood ratio function is given by

$$-2\log \widetilde{R}^{(2)} = 2\sum_{i=1}^{T} \log(1 + \lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(2)}), \tag{7}$$

where λ is the solution of the equation: $\sum_{i=1}^{T} \frac{\widetilde{\mathbf{z}}_i^{(2)}}{1+\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(2)}} = 0$.

Theorem 2.2. *Suppose that conditions (C1)-(C4) in Appendix hold. Then under the null hypothesis H_0, we have*

$$-2\log \widetilde{R}^{(2)} \xrightarrow{d} \chi_p^2.$$

Let $\widehat{Y}_i = \frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)} Y_i + \{1 - \frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)}\} f(\mathbf{x}_i, \widehat{\theta})$ be another imputation estimator of Y_i. Then the third EL ratio function is given as follows:

$$\widetilde{R}^{(3)} = \max\left\{\prod_{i=1}^{T} Tp_i \,\Big|\, \sum_{i=1}^{T} p_i \widetilde{\mathbf{z}}_i^{(3)} = 0, \sum_{i=1}^{T} p_i = 1, p_i \geq 0\right\} \tag{8}$$

where $\widetilde{\mathbf{z}}_i^{(3)} = \left(\widetilde{z}_{i1}^{(3)}, \widetilde{z}_{i2}^{(3)}, \cdots, \widetilde{z}_{ip}^{(3)}\right)^{\mathrm{T}}$, $\widetilde{z}_{ik}^{(3)} = \{\widehat{Y}_i - f(\mathbf{x}_i, \widehat{\theta})\} \cdot \{\widehat{Y}_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\}$, $i = 1, 2, \cdots, T$ and $k = 1, 2, \cdots, p$. Then the third empirical log-likelihood ratio function can be given by,

$$-2\log \widetilde{R}^{(3)} = 2\sum_{i=1}^{T} \log(1 + \lambda^{\mathrm{T}} \widetilde{\mathbf{z}}_i^{(3)}), \tag{9}$$

where λ is the solution of the equation: $\sum_{i=1}^{T} \frac{\widetilde{\mathbf{z}}_i^{(3)}}{1+\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(3)}} = 0$.

Theorem 2.3. *Suppose that conditions (C1)-(C4) in Appendix hold. Then under the null hypothesis H_0, we have*

$$-2\log \widetilde{R}^{(3)} \xrightarrow{d} \chi_p^2.$$

Remark 2.4. *For the sake of description, the above three methods based on Theorems 2.1-2.3 are denoted as IPW, IM1 and IM2, respectively. Although the ideas of constructing test statistics of IPW and IM2 are different, the empirical log-likelihood ratios for IPW and IM2 are consistent, since $\widetilde{z}_{ik}^{(3)} = \frac{\delta_i \delta_{i+k}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})}\{Y_i - f(\mathbf{x}_i, \widehat{\theta})\}\{Y_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\}$ which is equals to $\widetilde{z}_{ik}^{(1)}$.*

3. Simulations and Application

3.1. Simulation studies

In this section, we use several simulation examples to illustrate the finite sample performance of IPW, IM1, IM2 and CC methods for dealing with missing response data. The following simulated model is considered:

$$Y = \sin(\mathbf{x}^{\mathrm{T}}\theta) + (1 + \mathbf{x}^{\mathrm{T}}\theta)^2 + \varepsilon, \tag{10}$$

where $\theta = (0.5, 0.8)^{\mathrm{T}}$ and $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$ comes from binary normal distribution $\mathcal{N}(\mu, \Sigma)$ with $\mu = (0, 1)^{\mathrm{T}}$ and Σ = diag(1,4). To demonstrate the inference of different missing rates, the missing mechanisms are chosen from the following two logistic models:

$$\text{Case 1}: \ \pi_1(\mathbf{x}) = P(\delta = 1 \mid \mathbf{x} = x) = 1/[1 + \exp\{-(2.1 + 0.2x_1 + 0.5x_2)\}],$$
$$\text{Case 2}: \ \pi_2(\mathbf{x}) = P(\delta = 1 \mid \mathbf{x} = x) = 1/[1 + \exp\{-(0.8 + 0.8x_1 + 0.2x_2)\}].$$

The corresponding missing rates of Cases 1 and 2 are about 10% and 30%, respectively. The model errors ε_i obey the following two different models:

$$\text{AR(2)}: \varepsilon_i = a_1\varepsilon_{i-1} + a_2\varepsilon_{i-2} + e_i \ \text{ and } \ \text{MA(2)}: \varepsilon_i = a_1 e_{i-1} + a_2 e_{i-2} + e_i,$$

where e_i are generated from $N(0,1)$ and $U(-1,1)$, respectively. We take (a_1, a_2) = (0, 0), (0,0.4), (0.2,0.6), (-0.3,0.5), (0.5,-0.8). The power for each given (a_1, a_2) is evaluated among 1000 simulations for n = 50, 100, 150 and 200 at significance level $\alpha = 0.05$. The estimated sizes and powers of test are given in Tables 1-4, respectively for different settings.

From Tables 1-4, we can conclude that the estimated sizes and powers with different error distributions and missing rates are quite good. First, as the sample size increases, the estimated sizes converge toward their nominal size and the powers increase rapidly as (a_1, a_2) far away from (0,0). Second, the IM1 method give the best performance while the CC approach offers the worst. The IPW and IM2 methods give the same results which coincide with Remark 2.1. It is also interesting to note that for each method, the test with small missing rate performs better than that of large missing rate for fixed sample size and parameter vector (a_1, a_2). To show the performance of the proposed tests graphically, we consider $a_1 = 0$ and $a_2 = 0 : 0.1 : 1$. The power curves are depicted in Figures 1-2 for different scenarios, which show similar phenomenons described above.

Table 1: Estimated sizes and powers of four different test statistics for AR(2) with $e \sim N(0,1)$.

		$n=50$		$n=100$		$n=150$		$n=200$	
(a_1,a_2)	Method	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$
(0,0)	CC	0.103	0.156	0.076	0.093	0.064	0.082	0.057	0.078
	IPW	0.099	0.151	0.073	0.088	0.062	0.080	0.053	0.070
	IM1	0.093	0.126	0.065	0.073	0.058	0.071	0.061	0.056
	IM2	0.099	0.151	0.073	0.088	0.062	0.080	0.053	0.070
(0,0.4)	CC	0.581	0.405	0.865	0.600	0.957	0.762	0.991	0.862
	IPW	0.618	0.445	0.896	0.678	0.970	0.853	0.992	0.919
	IM1	0.622	0.441	0.899	0.703	0.973	0.891	0.995	0.938
	IM2	0.618	0.445	0.896	0.678	0.970	0.853	0.992	0.919
(0.2,0.6)	CC	0.916	0.733	0.998	0.947	1.000	0.994	1.000	0.998
	IPW	0.949	0.815	0.998	0.983	1.000	0.998	1.000	1.000
	IM1	0.947	0.819	1.000	0.992	1.000	1.000	1.000	1.000
	IM2	0.949	0.815	0.998	0.983	1.000	0.998	1.000	1.000
(-0.3,0.5)	CC	0.936	0.821	0.998	0.952	1.000	0.992	1.000	0.998
	IPW	0.953	0.907	0.998	0.983	1.000	1.000	1.000	1.000
	IM1	0.960	0.919	0.999	0.988	1.000	1.000	1.000	1.000
	IM2	0.953	0.907	0.998	0.983	1.000	1.000	1.000	1.000
(0.5,-0.8)	CC	0.990	0.865	1.000	0.983	1.000	0.999	1.000	1.000
	IPW	0.997	0.943	1.000	0.996	1.000	0.999	1.000	1.000
	IM1	0.997	0.948	1.000	0.998	1.000	1.000	1.000	1.000
	IM2	0.997	0.943	1.000	0.996	1.000	0.999	1.000	1.000

Table 2: Estimated sizes and powers of four different test statistics for AR(2) with $e \sim U(-1,1)$.

		$n=50$		$n=100$		$n=150$		$n=200$	
(a_1,a_2)	Method	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$
(0.0,0.0)	CC	0.087	0.128	0.057	0.073	0.060	0.060	0.059	0.068
	IPW	0.096	0.118	0.053	0.073	0.055	0.056	0.049	0.067
	IM1	0.084	0.104	0.057	0.061	0.058	0.064	0.580	0.065
	IM2	0.096	0.118	0.053	0.073	0.055	0.056	0.049	0.067
(0.0,0.4)	CC	0.562	0.368	0.879	0.622	0.971	0.772	0.993	0.874
	IPW	0.597	0.436	0.906	0.695	0.979	0.860	0.999	0.941
	IM1	0.613	0.442	0.919	0.735	0.981	0.892	0.999	0.963
	IM2	0.597	0.436	0.906	0.695	0.979	0.860	0.999	0.941
(0.2,0.6)	CC	0.926	0.734	1.000	0.951	1.000	0.994	1.000	1.000
	IPW	0.939	0.819	1.000	0.983	1.000	1.000	1.000	1.000
	IM1	0.940	0.828	1.000	0.989	1.000	1.000	1.000	1.000
	IM2	0.939	0.819	1.000	0.983	1.000	1.000	1.000	1.000
(-0.3,0.5)	CC	0.946	0.818	0.999	0.968	1.000	0.999	1.000	1.000
	IPW	0.957	0.897	1.000	0.991	1.000	0.999	1.000	1.000
	IM1	0.959	0.904	1.000	0.995	1.000	1.000	1.000	1.000
	IM2	0.957	0.897	1.000	0.991	1.000	0.999	1.000	1.000
(0.5,-0.8)	CC	0.995	0.882	1.000	0.986	1.000	1.000	1.000	1.000
	IPW	0.997	0.935	1.000	0.998	1.000	1.000	1.000	1.000
	IM1	0.998	0.943	1.000	1.000	1.000	1.000	1.000	1.000
	IM2	0.997	0.935	1.000	0.998	1.000	1.000	1.000	1.000

Table 3: Estimated sizes and powers of four different test statistics for MA(2) with $e \sim N(0,1)$.

		$n=50$		$n=100$		$n=150$		$n=200$	
(a_1,a_2)	Method	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$
(0.0,0.0)	CC	0.109	0.173	0.074	0.106	0.076	0.095	0.059	0.071
	IPW	0.103	0.136	0.071	0.100	0.071	0.087	0.049	0.069
	IM1	0.102	0.122	0.070	0.086	0.066	0.076	0.048	0.053
	IM2	0.103	0.136	0.071	0.100	0.071	0.087	0.049	0.069
(0.0,0.4)	CC	0.530	0.362	0.818	0.533	0.953	0.672	0.985	0.758
	IPW	0.557	0.405	0.849	0.618	0.970	0.751	0.990	0.852
	IM1	0.562	0.392	0.862	0.635	0.974	0.809	0.995	0.900
	IM2	0.557	0.405	0.849	0.618	0.970	0.751	0.990	0.852
(0.2,0.6)	CC	0.838	0.590	0.985	0.829	1.000	0.941	1.000	0.982
	IPW	0.864	0.660	0.991	0.895	1.000	0.972	1.000	0.990
	IM1	0.869	0.669	0.993	0.924	1.000	0.993	1.000	1.000
	IM2	0.864	0.660	0.991	0.895	1.000	0.972	1.000	0.990
(-0.3,0.5)	CC	0.646	0.426	0.907	0.572	0.987	0.748	0.999	0.861
	IPW	0.675	0.480	0.923	0.669	0.988	0.843	1.000	0.917
	IM1	0.671	0.490	0.936	0.712	0.996	0.888	1.000	0.959
	IM2	0.675	0.480	0.923	0.669	0.988	0.843	1.000	0.917
(0.5,-0.8)	CC	0.971	0.789	1.000	0.967	1.000	0.992	1.000	1.000
	IPW	0.981	0.856	1.000	0.991	1.000	0.999	1.000	1.000
	IM1	0.980	0.880	1.000	0.997	1.000	1.000	1.000	1.000
	IM2	0.981	0.856	1.000	0.991	1.000	0.999	1.000	1.000

Table 4: Estimated sizes and powers of four different test statistics for MA(2) with $e \sim U(-1,1)$.

		$n=50$		$n=100$		$n=150$		$n=200$	
(a_1,a_2)	Method	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$	$\pi_1(\mathbf{x})$	$\pi_2(\mathbf{x})$
(0.0,0.0)	CC	0.072	0.120	0.070	0.090	0.059	0.071	0.047	0.067
	IPW	0.070	0.114	0.055	0.079	0.058	0.068	0.050	0.062
	IM1	0.064	0.097	0.060	0.077	0.047	0.058	0.052	0.054
	IM2	0.070	0.114	0.055	0.079	0.058	0.068	0.050	0.062
(0.0,0.4)	CC	0.485	0.331	0.832	0.511	0.964	0.665	0.991	0.822
	IPW	0.517	0.357	0.868	0.582	0.970	0.771	0.995	0.879
	IM1	0.530	0.371	0.880	0.629	0.983	0.818	0.995	0.931
	IM2	0.517	0.357	0.868	0.582	0.970	0.771	0.995	0.879
(0.2,0.6)	CC	0.847	0.580	0.992	0.851	1.000	0.972	1.000	0.995
	IPW	0.886	0.662	0.996	0.920	1.000	0.981	1.000	0.999
	IM1	0.890	0.683	0.996	0.956	1.000	0.994	1.000	1.000
	IM2	0.886	0.662	0.996	0.920	1.000	0.981	1.000	0.999
(-0.3,0.5)	CC	0.622	0.377	0.952	0.629	0.992	0.779	1.000	0.895
	IPW	0.698	0.437	0.961	0.727	0.996	0.875	1.000	0.950
	IM1	0.702	0.457	0.970	0.780	0.997	0.917	1.000	0.984
	IM2	0.698	0.437	0.961	0.727	0.996	0.875	1.000	0.950
(0.5,-0.8)	CC	0.985	0.790	1.000	0.978	1.000	0.997	1.000	1.000
	IPW	0.991	0.885	1.000	0.992	1.000	1.000	1.000	1.000
	IM1	0.995	0.900	1.000	0.995	1.000	1.000	1.000	1.000
	IM2	0.991	0.885	1.000	0.992	1.000	1.000	1.000	1.000

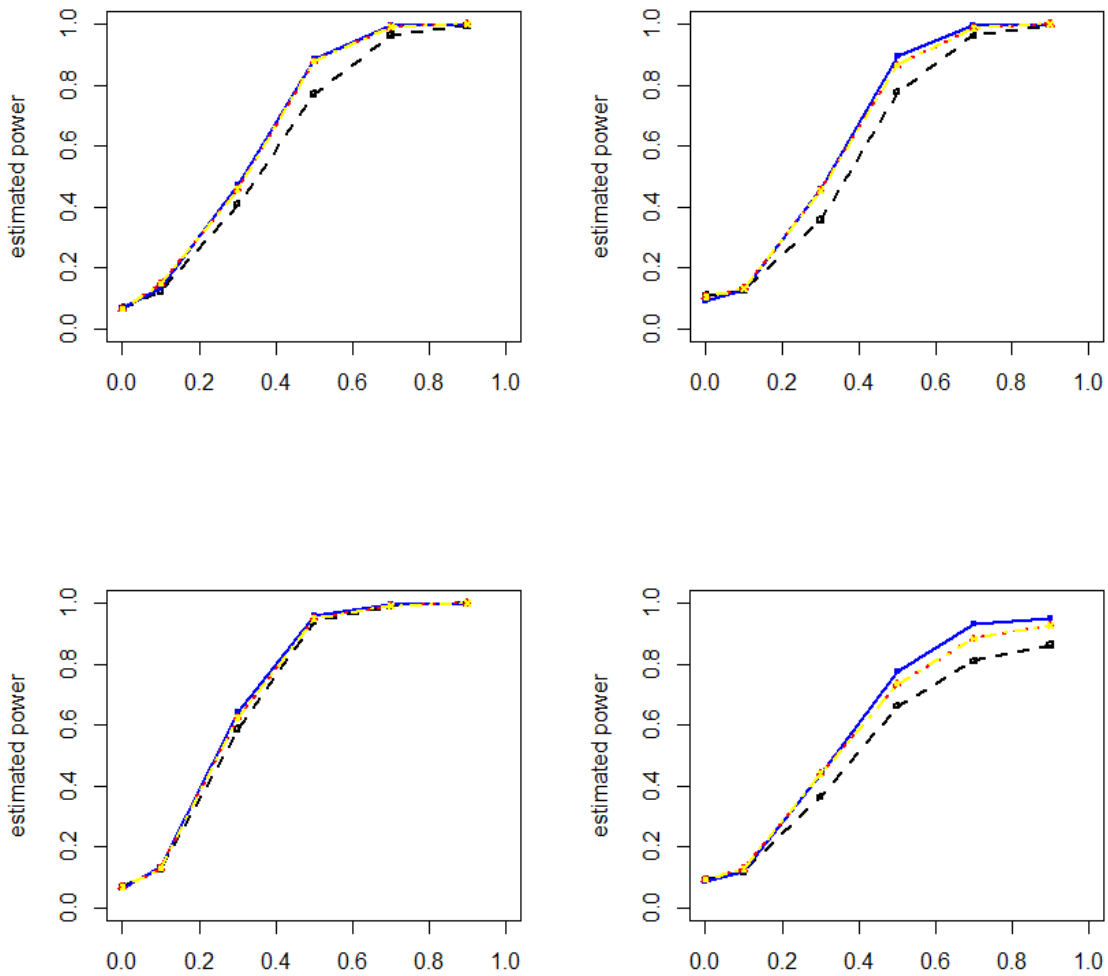

Figure 1: The power curves of tests based on IPW (dotted line), IM1 (solid line), IM2 (dot-dashed line) and CC (dashed line) methods for $e \sim N(0, 1)$ based on MR= 10% (left panel) and MR= 30% (right panel) with AR(2) (top) and MA(2) (bottom) under $a_1 = 0$ and different values of a_2 when $n = 100$.

3.2. *A real data example*

In this subsection, we apply the proposed test methods to analyze a real data collected from HIV clinical trials. The data set can be available in the R package "speff2trial". In the data set, the response Y stands for 1025 male patients who had received antiretroviral therapy prior to the trial. Based on the way of therapy, the data set can be divided into two subsets. The first data set is 253 male patients who applied monotherapy; the second data set is 772 male patients who adopted combined therapies. For each data set, Y is CD4 counts at 96 ± 5 weeks post therapy, X_1 is CD4 cell counts at baseline, X_2 is CD4 counts at 20 ± 5 weeks, X_3 is CD8 cell counts at baseline, X_4 is the CD8 cell counts at 20 ± 5 weeks. Due to death and dropout, there were 37.55% and 36.14% missing rates in the first and the second subset, respectively. The covariates X_1, X_2, X_3 and X_4 for all patients are available. The data set has been used by [4] to test

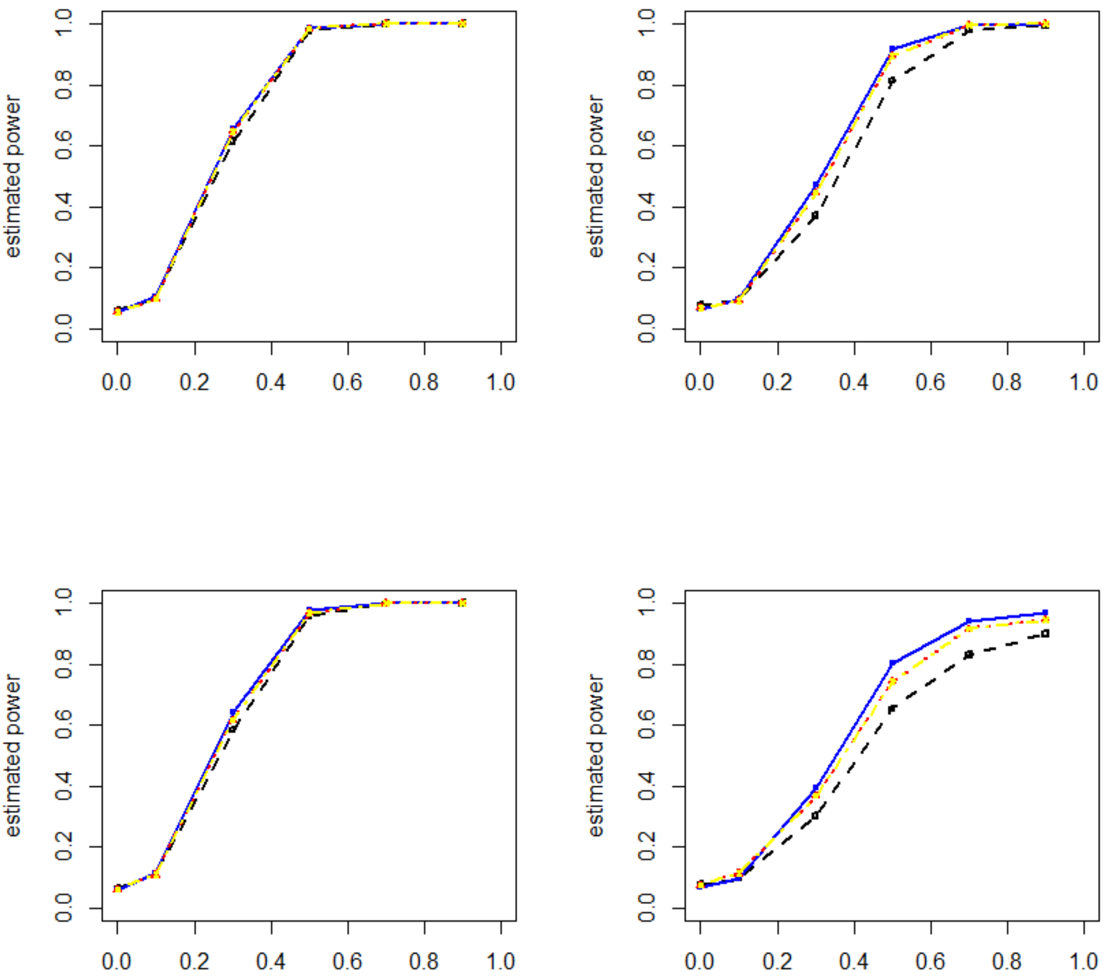

Figure 2: The power curves of tests based on IPW (dotted line), IM1 (solid line), IM2 (dot-dashed line) and CC (dashed line) methods for $e \sim U(-1,1)$ baed on MR= 10% (left panel) and MR= 30% (right panel) with AR(2) (top) and MA(2) (bottom) under $a_1 = 0$ and different values of a_2 when $n = 100$.

nonparametric component in partial linear model where the response is MAR. According to the conclusion of [4], these two subsets can be fitted with linear regression forms. We shall use these two subsets to test the serial correlation via linear regression forms. By the IPW and least squares methods, the fitted models are

$$\text{LM 1}: \ \widehat{Y} = 0.323X_1 + 0.684X_2 - 0.048X_3 - 0.017X_4,$$

$$\text{LM 2}: \ \widehat{Y} = 0.385X_1 + 0.608X_2 + 0.011X_3 - 0.051X_4.$$

Further, we test whether the residuals in LM1 and LM2 are first-order and second-order autocorrelated. The values of the test statistics based on the IPW, IM1, IM2 and CC methods are given in Table 5. We can see that the values of the test statistics with first-order and second-order autocorrelated are all less than

Table 5: The values of the test statistics based on HIV clinical trials.

	order	CC	IPW	IM1	IM2	critical value
LM1	1	0.149	0.283	0.149	0.283	3.841
	2	0.121	0.339	0.189	0.339	5.991
LM2	1	1.309	0.850	1.309	0.850	3.841
	2	0.086	1.931	2.452	1.931	5.991

Table 6: LB test results for the real data analysis.

	order	test statistic	p-value
LM1	1	0.1428	0.7055
	2	0.1662	0.9203
LM2	1	1.1058	0.2930
	2	2.2424	0.3259

3.841 and 5.991, respectively which are the 95% quantiles of central chi-square distribution with 1 and 2 degrees of freedom. Thus, it can be concluded that the residual sequence is not correlated. To further verify this conclusion, the Ljung-Box (LB) test for first-order and second-order autocorrelated is used and the corresponding testing results are displayed in Table 6, where the missing data are omitted naively. From Table 6, it can be seen that the p-values are all greater than the significance level 0.05, which indicates that the residuals are independent of each other. Hence, serial correlation does not exist in LM1 and LM2.

4. Concluding Remarks

In this paper, we applied the empirical likelihood approach to test the serial correlation for the residuals in parametric regression models with response missing at random. Three different empirical log-likelihood ratio test statistics based on IPW, IM1 and IM2 methods are proposed. The proposals can test not only zero first-order serial correlation, but also higher-order serial correlation. Besides, the computation is fast and easy to implement. The simulation study shows that IM1 method outperforms the other methods.

Appendix: Proof of Theorems

In this paper, the following conditions are required in order to obtain main conclusions. Denote $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ for any vector or matrix $\mathbf{A}$.

(C1) For any $\mathbf{x}$, $f(\mathbf{x}, \theta)$ is a continuous function of θ and the second derivatives with respect to θ are continuous, $\theta \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a compact set;

(C2) $E\left\{\frac{\partial f(\mathbf{x},\theta)}{\partial \theta}\right\}^{\otimes 2}$ is finite and positive definite;

(C3) The selection probability function $\pi(\mathbf{x})$ has a bounded continuous second derivatives almost surely and $\inf_{\mathbf{x}} \pi(\mathbf{x}) > 0$;

(C4) $\sup_x E(\varepsilon^4|\mathbf{x} = x) < \infty$ and $E\|\mathbf{x}\|^4 < \infty$.

Lemma 4.1. *Under conditions (C1)-(C4) and the null hypothesis, we have* $\widehat{\theta} - \theta = O_p\left(n^{-1/2}\right)$.

This lemma can be verified by Lemma 2 in [33].

Lemma 4.2. *Under the conditions (C1)-(C4) and the null hypothesis, we have* $T^{-1/2}\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)} \xrightarrow{d} N\left(\mathbf{0}, E\left(\frac{\sigma^4(\mathbf{x})}{\pi^2(\mathbf{x})}\right)\boldsymbol{I}_p\right)$, *where* $\boldsymbol{I}_p$ *is* $p\times p$ *identity matrix.*

Proof. Let $\widetilde{z}_{ik}^{(1)} = \frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)}\{Y_i - f(\mathbf{x}_i,\widehat{\theta})\}\frac{\delta_{i+k}}{\widehat{\pi}(\mathbf{x}_{i+k})}\{Y_{i+k} - f(\mathbf{x}_{i+k},\widehat{\theta})\}$, $\widetilde{\mathbf{z}}_i^{(1)} = (\widetilde{z}_{i1}^{(1)}, \widetilde{z}_{i2}^{(1)}, \cdots, \widetilde{z}_{ip}^{(1)})^{\mathrm{T}}$, $i = 1, 2, \cdots, T$, $k = 1, 2, \cdots, p$, $T = n - p$. Observe that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\widetilde{z}_{ik}^{(1)} =& \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)}\left\{Y_i - f(\mathbf{x}_i,\widehat{\theta})\right\}\frac{\delta_{i+k}}{\widehat{\pi}(\mathbf{x}_{i+k})}\left\{Y_{i+k} - f(\mathbf{x}_{i+k},\widehat{\theta})\right\}\\
=&\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})}\widehat{\varepsilon_i\varepsilon}_{i+k}\\
=&\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\widehat{\varepsilon_i\varepsilon}_{i+k} - \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}\{\widehat{\pi}(\mathbf{x}_{i+k}) - \pi(\mathbf{x}_{i+k})\}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})\pi(\mathbf{x}_{i+k})}\widehat{\varepsilon_i\varepsilon}_{i+k}\\
&-\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}\{\widehat{\pi}(\mathbf{x}_i) - \pi(\mathbf{x}_i)\}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})\pi(\mathbf{x}_i)}\widehat{\varepsilon_i\varepsilon}_{i+k}\\
&-\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}\{\widehat{\pi}(\mathbf{x}_i) - \pi(\mathbf{x}_i)\}\{\widehat{\pi}(\mathbf{x}_{i+k}) - \pi(\mathbf{x}_{i+k})\}}{\widehat{\pi}(\mathbf{x}_i)\widehat{\pi}(\mathbf{x}_{i+k})\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\widehat{\varepsilon_i\varepsilon}_{i+k}\\
:=&Z_{k1}^{(1)} - Z_{k2}^{(1)} - Z_{k3}^{(1)} - Z_{k4}^{(1)}.
\end{aligned}\tag{A.1}$$

For the term $Z_{k1}^{(1)}$ in (A.1), it can be further divided as

$$\begin{aligned}
Z_{k1}^{(1)} =& \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\widehat{\varepsilon_i\varepsilon}_{i+k} = \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\varepsilon_i\varepsilon_{i+k}\\
&-\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\varepsilon_i\left\{f(\mathbf{x}_{i+k},\widehat{\theta}) - f(\mathbf{x}_{i+k},\theta)\right\} - \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\varepsilon_{i+k}\left\{f(\mathbf{x}_i,\widehat{\theta}) - f(\mathbf{x}_i,\theta)\right\}\\
&+\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\left\{f(\mathbf{x}_i,\widehat{\theta}) - f(\mathbf{x}_i,\theta)\right\}\cdot\left\{f(\mathbf{x}_{i+k},\widehat{\theta}) - f(\mathbf{x}_{i+k},\theta)\right\}\\
:=&Z_{k1,1}^{(1)} - Z_{k1,2}^{(1)} - Z_{k1,3}^{(1)} + Z_{k1,4}^{(1)}.
\end{aligned}\tag{A.2}$$

For $Z_{k1,1}^{(1)}$ in (A.2), note that under null hypothesis $\mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\varepsilon_i\varepsilon_{i+k}\right\} = 0$, we find

$$\begin{aligned}
&\mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})}\varepsilon_i\varepsilon_{i+k}\right\}^2 = \mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi^2(\mathbf{x}_i)\pi^2(\mathbf{x}_{i+k})}\varepsilon_i^2\varepsilon_{i+k}^2\right\}\\
&= \mathrm{E}\left[\mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi^2(\mathbf{x}_i)\pi^2(\mathbf{x}_{i+k})}\varepsilon_i^2\varepsilon_{i+k}^2 \mid \mathbf{x}_i,\mathbf{x}_{i+k}\right\}\right] = \mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi^2(\mathbf{x}_i)\pi^2(\mathbf{x}_{i+k})}\mathrm{E}\left(\varepsilon_i^2\varepsilon_{i+k}^2 \mid \mathbf{x}_i,\mathbf{x}_{i+k}\right)\right\}\\
&= \mathrm{E}\left\{\frac{\delta_i\delta_{i+k}}{\pi^2(\mathbf{x}_i)\pi^2(\mathbf{x}_{i+k})}\sigma^2(\mathbf{x}_i)\sigma^2(\mathbf{x}_{i+\mathrm{k}})\right\} = \mathrm{E}\left\{\frac{\sigma^4(\mathbf{x})}{\pi^2(\mathbf{x})}\right\}.
\end{aligned}\tag{A.3}$$

From conditions (C1) and (C2), we can derive that

$$\begin{aligned} Z_{k1,2}^{(1)} =& \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \frac{\delta_i \delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})} \varepsilon_i \left\{ f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \theta) \right\} \\ =& \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \frac{\delta_i \delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})} \varepsilon_i \frac{\partial f(\mathbf{x}_{i+k}, \theta)}{\partial \theta^{\tau}} (\widehat{\theta} - \theta) + (\widehat{\theta} - \theta)^{\tau} \frac{1}{T} \sum_{i=1}^{T} \frac{\delta_i \delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})} \varepsilon_i \frac{\partial^2 f(\mathbf{x}_{i+k}, \tilde{\theta})}{\partial \theta \partial \theta^{\tau}} (\widehat{\theta} - \theta) \\ :=& R_{k1,1}^{(1)} (\widehat{\theta} - \theta) + (\widehat{\theta} - \theta)^{\tau} R_{k1,2}^{(1)} (\widehat{\theta} - \theta), \end{aligned} \tag{A.4}$$

where $\tilde{\theta}$ lies between $\widehat{\theta}$ and θ. It is easy to gain $R_{k1,1}^{(1)} = O_p\left(T^{-\frac{1}{2}}\right)$. According to Lemma 4.1 and the continuity of $\partial^2 f\left(\mathbf{x}_i, \theta\right) / \partial \theta \partial \theta^{\tau}$ as a function of θ, we can obtain that

$$E \left| \frac{\delta_i \delta_{i+k}}{\pi\left(\mathbf{x}_i\right) \pi\left(\mathbf{x}_{i+k}\right)} \varepsilon_i \frac{\partial^2 f(\mathbf{x}_{i+k}, \tilde{\theta})}{\partial \theta \partial \theta^{\tau}} \right| = E \left\{ \frac{\delta_i \delta_{i+k}}{\pi\left(\mathbf{x}_i\right) \pi\left(\mathbf{x}_{i+k}\right)} \left|\varepsilon_i\right| \left| \frac{\partial^2 f\left(\mathbf{x}_{i+k}, \theta\right)}{\partial \theta \partial \theta^{\tau}} \right| \right\} = O(1).$$

Therefore $R_{k1,2}^{(1)} = O_p(1)$. Then we can conclude that $Z_{k1,2}^{(1)} = o_p(1)$. Similarly to the derivation for $Z_{k1,2}^{(1)}$, we can derive that $Z_{k1,3}^{(1)} = o_p(1)$. Note that $f(\mathbf{x}_i, \widehat{\theta}) - f(\mathbf{x}_i, \theta) = \frac{\partial f(\mathbf{x}_i, \widetilde{\theta}_1)}{\partial \theta} (\widehat{\theta} - \theta)$, where $\widetilde{\theta}_1$ lies between $\widehat{\theta}$ and θ, $f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \theta) = \frac{\partial f(\mathbf{x}_{i+k}, \tilde{\theta}_2)}{\partial \theta^{\tau}} (\widehat{\theta} - \theta)^{\tau}$, where $\widetilde{\theta}_2$ lies between $\widehat{\theta}$ and θ. Then, we have

$$\begin{aligned} Z_{k1,4}^{(1)} &= (\widehat{\theta} - \theta)^{\tau} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \left\{ \frac{\delta_i \delta_{i+k}}{\pi(\mathbf{x}_i)\pi(\mathbf{x}_{i+k})} \cdot \frac{\partial f(\mathbf{x}_i, \tilde{\theta}_1)}{\partial \theta} \cdot \frac{\partial f(\mathbf{x}_{i+k}, \tilde{\theta}_2)}{\partial \theta^{\tau}} \right\} (\widehat{\theta} - \theta) \\ &\stackrel{\text{def}}{=} (\widehat{\theta} - \theta)^{\tau} R_{k1,3}^{(1)} (\widehat{\theta} - \theta). \end{aligned}$$

Similar to the argument for $R_{k1,2}^{(1)}$, we can get $R_{k1,3}^{(1)} = O_p(1)$ and hence, $Z_{k1,4}^{(1)} = o_p(1)$. Then it follows that

$$Z_{k1}^{(1)} = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \frac{\delta_i \delta_{i+k}}{\pi\left(\mathbf{x}_i\right) \pi\left(\mathbf{x}_{i+k}\right)} \varepsilon_i \varepsilon_{i+k} + o_p(1) \stackrel{\text{def}}{=} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} Z_{k11,i} + o_p(1). \tag{A.5}$$

Similarly, we can obtain $Z_{km}^{(1)} = o_p(1)$, for $m = 2, 3, 4$. Denote $\mathbf{z}_{ik}^{(1)} = (Z_{111,i}, Z_{211,i}, \ldots, Z_{p11,i})^{\tau}$. Then by (A.5), we have $\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \widetilde{\mathbf{z}}_i^{(1)} = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \mathbf{z}_i^{(1)} + o_p(1)$. Assuming that $\boldsymbol{\varphi}$ is any p-dimension nonzero vector, then $\boldsymbol{\varphi}^{\tau} \mathbf{z}_i^{(1)}$ is a p-dependent random variable sequence under null hypothesis and $\mathrm{Cov}\left(\boldsymbol{\varphi}^{\tau} \mathbf{z}_{k1,1}^{(1)}, \boldsymbol{\varphi}^{\tau} \mathbf{z}_{k1,1}^{(1)}\right) = \boldsymbol{\varphi}^{\tau} \mathrm{Cov}\left(\mathbf{z}_i^{(1)}, \mathbf{z}_j^{(1)}\right) \boldsymbol{\varphi} = 0$, for $i \neq j$. Then, according to the central limit theorem for the p-dependent sequence of [23], $\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{\varphi}^{\tau} \mathbf{z}_i^{(1)} \xrightarrow{d} N(0, \Omega)$, where $\Omega = \boldsymbol{\varphi}^{\tau} \boldsymbol{\varphi} E\left\{\frac{\sigma^4(\mathbf{x})}{\pi^2(\mathbf{x})}\right\}$. According to Cramer-Wold device, we obtain that $\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \widetilde{\mathbf{z}}_i^{(1)} \xrightarrow{d} N\left(\mathbf{0}, E\left(\frac{\sigma^4(\mathbf{x})}{\pi^2(\mathbf{x})}\right) \boldsymbol{I}_p\right)$, which completes the proof of Lemma 4.2.

Lemma 4.3. *Under the conditions (C1)-(C4) and the null hypothesis, we have*

$$\frac{1}{T} \sum_{i=1}^{T} \widetilde{\mathbf{z}}_i^{(1)} \widetilde{\mathbf{z}}_i^{(1)^T} \xrightarrow{p} E\left(\frac{\sigma^4(\mathbf{x})}{\pi^2(\mathbf{x})}\right) \boldsymbol{I}_p.$$

Proof. Similar to the proofs of Lemma 4.2, we can show Lemma 4.3 holds easily.

Appendix B: Proof of Theorem 2.1

Let $\widetilde{\mathbf{z}}^{\star} \stackrel{\text{def}}{=} \max_{1 \le i \le T} \left\| \widetilde{\mathbf{z}}_i \right\|$. By Lemmas 4.2 and 4.3 and the same arguments of Lemma 5.6 in [10], we can derive that $\widetilde{\mathbf{z}}^{\star} = o_p\left(T^{1/2}\right)$ and $\|\lambda\| = O_p(T^{-1/2})$. Then, applying a Taylor expansion to (5), the empirical log-likelihood ratio function (5) can be decomposed as follows,

$$-2\log \widetilde{R}^{(1)} = 2\sum_{i=1}^{T}\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)} - \sum_{i=1}^{T}\left(\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)}\right)^2 + o_p(1). \tag{A.6}$$

By some calculations, we can obtain

$$\sum_{i=1}^{T}\left\{\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)^{\mathrm{T}}}\right\}^2 = \sum_{i=1}^{T}\lambda^{\mathrm{T}}\widetilde{\mathbf{z}}_i^{(1)^{\mathrm{T}}} + o_p(1) \text{ and } \lambda = \left\{\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)}\widetilde{\mathbf{z}}_i^{(1)^{\mathrm{T}}}\right\}^{-1}\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)} + o_p(T^{-1/2}).$$

Thus, by (A.6), we have

$$-2\log \widetilde{R}^{(1)} = \left\{\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)}\right\}^{\mathrm{T}}\left\{\frac{1}{T}\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)}\widetilde{\mathbf{z}}_i^{(1)^{\mathrm{T}}}\right\}^{-1}\left\{\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\widetilde{\mathbf{z}}_i^{(1)}\right\} + o_p(1).$$

Then according to Lemmas 4.2-4.3 and Slutsky's Theorem, we can get $-2\log \widetilde{R}^{(1)} \xrightarrow{d} \chi_p^2$.

Appendix C: Proof of Theorem 2.2

Let $\widetilde{z}_{ik}^{(2)} = \{\delta_i Y_i + (1-\delta_i) f(\mathbf{x}_i, \widehat{\theta}) - f(\mathbf{x}_i, \widehat{\theta})\}\cdot\{\delta_{i+k}Y_{i+k} + (1-\delta_{i+k}) f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \widehat{\theta})\}$, $\widetilde{\mathbf{z}}_{IM}^{(2)} = (\widetilde{z}_{i1}^{(2)}, \widetilde{z}_{i2}^{(2)}, \cdots, \widetilde{z}_{ip}^{(2)})^{\mathrm{T}}$, $i = 1, 2, \cdots, T$, $T = n - p$. Observe that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\tilde{z}_{ik}^{(2)} &= \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\left\{Y_i - f(\mathbf{x}_i, \widehat{\theta})\right\}\cdot\delta_{i+k}\left\{Y_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\right\}\\
&= \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\left\{f(\mathbf{x}_i, \theta) + \varepsilon_i - f(\mathbf{x}_i, \widehat{\theta})\right\}\cdot\delta_{i+k}\left\{f(\mathbf{x}_{i+k}, \theta) + \varepsilon_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\right\}\\
&= \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\delta_{i+k}\varepsilon_i\varepsilon_{i+k} - \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\delta_{i+k}\varepsilon_i\left\{f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \theta)\right\}\\
&\quad - \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\delta_{i+k}\varepsilon_{i+k}\left\{f(\mathbf{x}_i, \widehat{\theta}) - f(\mathbf{x}_i, \theta)\right\}\\
&\quad + \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\delta_i\delta_{i+k}\left\{f(\mathbf{x}_i, \widehat{\theta}) - f(\mathbf{x}_i, \theta)\right\}\left\{f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \theta)\right\}\\
&\overset{\text{def}}{=} \mathbf{z}_{k1}^{(2)} - \mathbf{z}_{k2}^{(2)} - \mathbf{z}_{k3}^{(2)} + \mathbf{z}_{k4}^{(2)}.
\end{aligned}$$

Similar to the proof of Lemma 4.2, we can obtain that $\mathbf{z}_{km}^{(2)} = o_p(1)$ for $m = 2, 3, 4$. Besides, it can be concluded that $\mathrm{E}\left(\delta_i\delta_{i+k}\varepsilon_i\varepsilon_{i+k}\right)^2 = E\left(\delta_i\delta_{i+k}\varepsilon_i^2\varepsilon_{i+k}^2\right) = \mathrm{E}\left[\mathrm{E}\left\{\delta_i\delta_{i+k}\varepsilon_i^2\varepsilon_{i+k}^2 \mid \mathbf{x}_i, \mathbf{x}_{i+k}\right\}\right] = \mathrm{E}\left\{\delta_i\delta_{i+k}\mathrm{E}\left(\varepsilon_i^2\varepsilon_{i+k}^2 \mid \mathbf{x}_i, \mathbf{x}_{i+k}\right)\right\} = \mathrm{E}\left\{\delta_i\delta_{i+k}\sigma^2(\mathbf{x}_i)\,\sigma^2(\mathbf{x}_{i+k})\right\} = \mathrm{E}\{\pi^2(\mathbf{x})\sigma^4(\mathbf{x})\}$. Then by the central limit theorem for the p-dependent sequence of [23] and Cramer-Wold's device, we derive that $\mathbf{z}_{k1}^{(2)} \xrightarrow{d} N\left(\mathbf{0}, E^2\{\pi(\mathbf{x})\sigma^2(\mathbf{x})\}\boldsymbol{I}_p\right)$, which together with the Slutsky's Theorem gives the result of Theorem 2.

Appendix D: Proof of Theorem 2.3

Observe that

$$\begin{aligned}\widetilde{z}_{ik}^{(3)} &= \{\widehat{Y}_i - f(\mathbf{x}_i, \widehat{\theta})\}\{\widehat{Y}_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\} = \Big[\frac{\delta_i}{\pi(\mathbf{x}_i)} Y_i + \Big\{1 - \frac{\delta_i}{\pi(\mathbf{x}_i)}\Big\} f(\mathbf{x}_i, \widehat{\theta}) - f(\mathbf{x}_i, \widehat{\theta})\Big] \\ &\quad \cdot \Big[\frac{\delta_{i+k}}{\pi(\mathbf{x}_{i+k})} Y_{i+k} + \Big\{1 - \frac{\delta_{i+k}}{\pi(\mathbf{x}_{i+k})}\Big\} f(\mathbf{x}_{i+k}, \widehat{\theta}) - f(\mathbf{x}_{i+k}, \widehat{\theta})\Big] \\ &= \frac{\delta_i}{\widehat{\pi}(\mathbf{x}_i)}\{Y_i - f(\mathbf{x}_i, \widehat{\theta})\} \frac{\delta_{i+k}}{\widehat{\pi}(\mathbf{x}_{i+k})}\{Y_{i+k} - f(\mathbf{x}_{i+k}, \widehat{\theta})\} = \widetilde{z}_{ik}^{(1)}.\end{aligned}$$

Then the result of Theorem 2.3 can be verified by Theorem 2.1.

References

[1] A. B. Owen, *Empirical likelihood ratio confidence intervals for a single functional*, Biometrika, **75(2)** (1988), 237–249.
[2] A. B. Owen, *Empirical likelihood ratio confidence regions*, The annals of statistics, **18(1)** (1990), 90–120.
[3] B. B. Chen, G. M. Pan, Q. Yang, W. Zhou, *Large dimensional empirical likelihood*, Statistica Sinica, **25** (2015), 1659–1677.
[4] C. Z. Niu, X. Guo, W. L. Xu, L. X. Zhu, *Checking nonparametric component for partial linear regression model with missing response*, Journal of Statistical Planning and Inference, **168**(2016), 1–19.
[5] D. Li, T. Stengos, *Testing serial correlation in semiparametric time series models*, Journal of Time Series Analysis, **24** (2003), 311–335.
[6] E. M. Chi, G. C. Reinsel, *Asymptotic properties of the score test for autocorrelation in a random effects with AR(1) errors model*, Statistics and Probability Letters, **11(5)** (1991), 453–457.
[7] E. M. Chi, G. C. Reinsel, *Models for longitudinal data with random effects and AR(1) Errors*, Journal of the American Statistical Association, **84(406)** (1989), 452–459.
[8] F. Liu, M. Chen, J. Z. Zou, *Empirical likelihood ratio tests for sequence correlation in partial linear models*, Acta Mathematicae Applicatae Sinica, **29(04)** (2006), 577–586.
[9] G. Dai, U. U. Müller, *Efficient estimators for expectations in nonlinear parametric regression models with responses missing at random*, Electronic Journal of Statistics, **13(2)** (2019), 3985–4014.
[10] G. L. Fan, H. X. Xu, H. Y. Liang, *Empirical likelihood inference for partially time-varying coefficient errors-in-variables models*, Electronic Journal of Statistics, **6** (2012), 1040–1058.
[11] G. L. Fan, H. X. Xu, H. Y. Liang, *Dimension reduction estimation for central mean subspace with missing multivariate response*, Journal of Multivariate Analysis, **174** (2019), 104542.
[12] H. Tsay, K. S. Chan, *Testing for nonlinearity with partially observed time series*, Biometrika, **87** (2000), 805–821.
[13] I. Van Keilegom, W. González Manteiga, C. Sánchez Sellero, *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution*, Test, **17(2)** (2008), 401–415.
[14] J. Alejo, G. Montes-Rojas, W. Sosa-Escudero, *Testing for serial correlation in hierarchical linear models*, Journal of Multivariate Analysis, **165** (2018), 101–116.
[15] J. Fan, L. S. Huang, *Goodness-of-fit tests for parametric regression models*, Journal of the American Statistical Association, **96(454)** (2001), 640–652.
[16] J. G. Ibrahim, *Incomplete data in generalized linear models*, Journal of the American Statistical Association, **85(411)** (1990), 765–769.
[17] J. M. Robins, A. Rotnitzky, L. P. Zhao, *Estimation of regression coefficients when some regressors are not always observed*, Journal of the American Statistical Association, **89(427)** (1994), 846–866.
[18] J. N. K. Rao, *On variance estimation with imputed survey data*, Journal of the American Statistical Association, **91(434)** (1996), 499–506.
[19] J. Shi, T. S. Lau, *Empirical likelihood for partially linear models*, Journal of Multivariate Analysis, **72** (2000), 132–148.
[20] J. X. Zheng, *A consistent test of functional form via nonparametric estimation techniques*, Journal of Econometrics, **75** (1996), 263–289.
[21] J. Yang, C. Wei, *Testing serial correlation in partially linear additive models*, Acta Mathematicae Applicatae Sinica, English Series, **35(2)** (2019), 401–411.
[22] P. E. Cheng, *Nonparametric estimation of mean functionals with data missing at random*, Journal of the American Statistical Association, **89(425)** (1994), 81–87.
[23] P. J. Brockwell, R. A. Davis, *Times series in theory and methods*, (2nd ed), Springer-Verlag, New York, 1991.
[24] Q. H. Wang, B. Y. Jing, *Empirical likelihood for partial linear models*, Annals of the Institute of Statistical Mathematics, **55(3)** (2003), 585–595.
[25] Q. Li, C. Hsiao, *Testing serial correlation in semiparametric panel data models*, Journal of Econometrics, **87(2)** (1998), 207–237.
[26] Q. Wang, O. Linton, W. Härdle, *Semiparametric regression analysis with missing response at random*, Journal of the American Statistical Association, **99(466)** (2004), 334–345.
[27] R. J. A. Little, D. B. Rubin, *Statistical analysis with missing data*, Wiley, New York, 1987.
[28] W. González-Manteiga, A. Peréz-González, *Goodness-of-fit tests for linear regression models with missing response data*, Canadian Journal of Statistics, **34(1)** (2006), 149–170.
[29] W. L. Xu, X. Guo, L. X. Zhu, *Goodness-of-fitting for partial linear model with missing response at random*, Journal of Nonparametric Statistics, **24(1)** (2012), 103–118.

[30] W. Xu, L. Zhu, *Testing the adequacy of varying coefficient models with missing responses at random*, Journal of the American Statistical Association, **76(1)** (2013), 53–69.

[31] W. Xu, X. Guo, *Nonparametric checks for varying coefficient models with missing response at random*, International Journal for Theoretical and Applied Statistics, **76(4)** (2013), 459–482.

[32] X. Guo, W. Xu, *Goodness-of-fit tests for general linear models with covariates missed at random*, Journal of Statistical Planning and Inference, **142(7)** (2012), 2047–2058.

[33] X. Guo, W. Xu, L. Zhu, *Model checking for parametric regressions with response missing at random*, Annals of the Institute of Statistical Mathematics, **67(2)** (2015), 229-259.

[34] X. H. Liu, Z. Z. Wang, X. M. Hu, G. F. Wang, *Testing serial correlation in partially linear single-index errors-in-variables models*, Communications in Statistics-Theory and Methods, **40(14)** (2011), 2554–2573.

[35] X. M. Hu, Z. Z. Wang, F. Liu, *Testing serial correlation in semiparametric varying-coefficient partially linear models*, Communications in Statistics-Theory and Methods, **38(13)** (2009), 2145–2163.

[36] Y. C. Qi, Y. C. Zhou, *Empirical likelihood method for complete independence test on high-dimensional data*, Journal of Statistical Computation and Simulation, **92(11)** (2022), 2386–2402.

[37] Y. Liu, B. Wei, *Testing of correlation and heteroscedasticity in nonlinear regression models with DBL(p, q, 1) random errors*, Acta Mathematica Scientia, **28B** (2008), 613–632.

[38] Y. Qin, T. Qiu, Q. Lei, *Empirical likelihood for response differences in two linear regression models with missing data*, Acta Mathematicae Applicatae Sinica, **31(04)** (2015), 963–976.

[39] Y. Y. Zou, C. X. Wu, G. L. Fan, R. Q. Zhang, *Estimation for a hybrid model of functional and linear measurement errors regression with missing response*, Statistics, **56** (2022), 271–296.

[40] Y. Zhou, A. T. K. Wan, X. Wang, *Estimating equations inference with missing data*, Journal of the American Statistical Association, **103(483)** (2008), 1187–1199.

[41] Z. E. Sun, S. J. Wang, *Semiparametric estimation in regression with missing covariates using single-index models*, Annals of the Institute of Statistical Mathematics, **71(5)** (2019), 1201–1232.

[42] Z. G. Zhou, W. M. Qian, C. He, *Testing serial correlation for partially nonlinear models*, Journal of the Korean Statistical Society, **39(4)** (2010), 501–509.

[43] Z. H. Sun, Q. H. Wang, *Checking the adequacy of a general linear model with responses missing at random*, Journal of Statistical Planning and Inference, **139** (2009), 3588–3604.