Empirical likelihood confidence regions for the parameters of a two phases nonlinear model with and without missing response data

Zahraa SALLOUM

*Université de Lyon, Université Lyon 1, CNRS, UMR 5208, Institut Camille Jordan, Bat. Braconnier, 43, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France

Abstract

In this paper, we use the empirical likelihood method to construct the confidence regions for the difference between the parameters of a two-phases nonlinear model. Two-phases nonlinear model with response variables missing at randoms is also studied by proposing three empirical likelihood statistics and it is shown that all this statistics have asymptotically chi-squared distributions. By Monte-Carlo simulations we show the performance of the proposed test statistics.

Keywords: Empirical likelihood, Confidence region, Two phases, Missing response.

AMS 2000 subject classifications : Primary 62J02; secondary 62F05.

1. Introduction

Let us consider the following nonlinear model

\[ Y_i = \begin{cases} 
    f(X_i; \beta) + \varepsilon_i & i = 1, \cdots, k, \\
    f(X_i; \beta_1) + \varepsilon_i & i = k+1, \cdots, n, 
\end{cases} \tag{1} \]

where \( \beta \) and \( \beta_1 \) are \( d \times 1 \) vectors of unknown parameters, \( X_i \) is a \( (p \times 1) \) random vector of regressors with distribution function \( H(x) \), with \( x \in \Upsilon \), where \( \Upsilon \subseteq \mathbb{R}^p \) is a compact set. Let us consider the vector \( Y = (Y_1, \cdots, Y_n) \), where, for each observation \( i \), \( Y_i \) denotes the response variable (which can have missing value) and \( \varepsilon_i \) is the error. The continuous random vector sequence \( (X_i, \varepsilon_i)_{1 \leq i \leq n} \) is independent identically distributed (i.i.d), with the same joint distribution as \( (X, \varepsilon) \). For all \( i \), \( \varepsilon_i \) is independent of \( X_i \).

Let us denote \( \gamma = \beta - \beta_1 \).

There are two aims in this paper. First, we suppose that, the response variable

\*Corresponding author

Email address: salloum@math.univ-lyon1.fr (Zahraa SALLOUM)
Y_i is observed for each observation i and we construct the confidence regions for γ in nonlinear model, or we test the null hypothesis

$$H_0 : \gamma = \gamma_0,$$

with γ_0 a known vector.

Second, we construct the confidence regions for γ, or we test H_0 when some values of Y may be missing and X_i is observed completely. That is, we obtain an incomplete sample \{(X_l, Y_l, δ_l)\}_{1 ≤ l ≤ n} from model (1), where all the X_l is observed, (δ_l)_{1 ≤ l ≤ n} is a sequence of random variables, such that δ_l = 0 if Y_l is missing and δ_l = 1 otherwise. We assume that Y_l is missing at random (MAR).

The MAR assumption implies that δ_l and Y_l are conditionally independent given X_l. That is, \(P[\delta_l = 1 | X_l, Y_l] = P[\delta_l = 1 | X_l],\) for all 1 ≤ l ≤ n. The MAR assumption is a common condition for statistical analysis with missing data and is reasonable in many practical situations, see Ciuperca [7], Little and Rubin [13] and Qin et al. [21].

In this kind of problem, we can use the bootstrap approach to construct confidence regions for γ, but, one of the inconvenience of the bootstrap is that, it needs some subjective instructions on the shapes and orientations of the confidence regions. In this paper, we will apply the empirical likelihood method for constructing the confidence regions nonparametrically, as an alternative to the bootstrap method. An important characteristic of empirical likelihood is that, it uses only the data to determine the shape and orientation of a confidence regions. This method was introduced by Owen ([18],[19]) as a way to extend the ideas of likelihood based inference to certain nonparametric situations.

Various authors extend empirical likelihood methodology to many statistical situations. Ciuperca and Salloum [8], Kim and Siegmund [11] and Liu et al. [15] used the empirical likelihood to detect the change-point in the regression parameters of the linear and nonlinear model. For a epidemic change model, Ning et al. [17] proposed a method based on the empirical likelihood to detect the epidemic changes of the mean after unknown change points. Kolaczyk [12] shows that empirical likelihood is justified as a method of inference for a class of linear models, and shows in particular how empirical likelihood may be used with generalized linear models. To construct the confidence regions for the coefficients in the linear regression model, Chen [4] proposed a nonparametric method based on empirical likelihood. Ciuperca [7], Qin et al. [21] and Xue [25] considered this same problem but for the models with missing response data. Always using the empirical likelihood method, Zi et al. [26] construct the confidence regions for the difference in value between coefficients of two-sample linear regression model with complete data and Wei et al. [24] for a model with missing response data.

In this paper, for the model (1), we use the empirical likelihood method to construct the confidence regions for the parameter \(\gamma = \beta - \beta_1,\) firstly, if the
response variable $Y_i$ is observed for each $i = 1, \cdots, n$, next when the response variable $Y_i$ can be missing. In each case, we give the empirical likelihood and confidence regions for $\gamma$ and we prove that all empirical likelihood statistics have a chi-squared asymptotic distribution. Then, we generalize the papers of Wei et al. [24] and of Zi et al. [26] in the nonlinear model case. One of the major difficulties for nonlinear model (beside the linear model approach) is that, for finding the test statistic, the corresponding score functions depend on the regression parameters, and above all, the analytical form of these derivatives is unknown. On the other hand, in the linear models, many proofs are based on the convexity of the regression function with respect to the parameter regression, then, the extreme value of a convex function is attained on the boundary. These two factors lead to a more difficult theoretical study of the test statistics for nonlinear model.

The paper is structured as follows. In Section 2, we introduce assumptions, some notations, null and alternative hypothesis. In Section 3, we construct the empirical likelihood ratio and the confidence regions of $\gamma$ for the model with complete data. The confidence regions of $\gamma$ for the model with missing response data corresponding to three empirical likelihood statistics are given in Section 4. In Section 5, simulations results illustrate the performance of the proposed empirical likelihood confidence regions in the two cases. Proofs of the main results and lemmas are given in Section 6.

2. Hypothesis, notations, assumptions

All vectors are column and $v^t$ denotes the transposed of $v$. All vectors and matrices are in bold. Concerning the used norms, for a m-vector $v = (v_1, \cdots, v_m)$, let us denote by $\|v\|_1 = \sum_{j=1}^m |v_j|$ its $L_1$-norm and $\|v\|_2 = \left(\sum_{j=1}^m v_j^2\right)^{1/2}$ its $L_2$-norm. For a matrix $D = (a_{ij})_{1 \leq i \leq m_1}^{1 \leq j \leq m_2}$, we denote by $\|D\|_1 = \max_{1 \leq j \leq m_2} \left(\sum_{i=1}^{m_1} |a_{ij}|\right)$, the subordinate norm to the vector norm $\|\cdot\|_1$. Let $\xrightarrow{L^p} \xrightarrow{P} \xrightarrow{a.s.}$ represent convergence in distribution, in probability and almost sure, respectively, as $n \to \infty$.

All throughout the paper, $C$ denotes a positive generic constant which may take different values in different formula or even in different parts of the same formula. Moreover, $\mathbf{0}_d$ and $\mathbf{1}_d$ denote the $d$-vectors with all components zero and 1, respectively.

For the model (1), the regression function $f : \Upsilon \times \Gamma \to \mathbb{R}$, with $\Upsilon \subseteq \mathbb{R}^p$ and $\Gamma \subseteq \mathbb{R}^d$, is known up to a $d$-dimensional parameter $\beta = (\beta_1, \cdots, \beta_d)$. The sets $\Upsilon$ and $\Gamma$ are compact.

With regard to the random variable $\varepsilon$ we make following assumption:

(A1) $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] < \infty$, for all $i = 1, \cdots, n$.

The regression function $f : \Upsilon \times \Gamma \to \mathbb{R}$ and the random vector $X$ satisfy the
conditions:

(A2) for all \( x \in \mathcal{Y} \) and for \( \beta \in \Gamma \), the function \( f(x, \beta) \) is thrice differentiable in \( \beta \) and continuous on \( \mathcal{Y} \).

In following, for \( x \in \mathcal{Y} \) and \( \beta \in \Gamma \), we use notation \( \dot{f}(x, \beta) \equiv \partial^2 f(x, \beta)/\partial \beta^2 \). 

(A3) \( (|\partial_j^2 f(x, \beta)|)_{1 \leq j, k \leq d} \) and \( (|\partial_j^2 f(x, \beta)|)_{1 \leq j, k, l \leq d} \) are bounded for any \( x \in \mathcal{Y} \) and \( \beta \) in a neighborhood of \( \beta^0 \).

(A4) \( E[|f(X, \beta)|_1] < \infty \), \( E[|f(X, \beta)\dot{f}(X, \beta)||] < \infty \) and \( E[|\partial^2 f(X, \beta)|] < \infty \), for all \( 1 \leq j, k \leq d \) and \( \beta \) in a neighborhood of \( \beta^0 \).

The assumptions (A3) and (A4) are standard conditions, which are used in nonlinear models, see the paper of Ciuperca [7] and the book of Seber and Wild [22] for example.

We will construct the confidence region for \( \gamma = \beta - \beta_1 \). At the same time, we test the hypothesis

\[ H_0 : \gamma = \gamma_0, \] (2)

where \( \gamma_0 \) is a \((d \times 1)\) known vector. The alternative hypothesis, is

\[ H_1 : \gamma \neq \gamma_0. \] (3)

Under \( H_0 \), let \( \beta^0 \) denote the true value of \( \beta \), where \( \beta \) is the generic value of the regression parameter for the first phase.

Consider the following sets \( I \equiv \{1, \ldots, k\} \) and \( J \equiv \{k + 1, \ldots, n\} \), which contain the observation subscripts of the two segments for the model \([1]\).

For \( i \in I \), let us consider the following \( d \)-random vectors

\[ g(X_i, \beta) \equiv g_i(\beta) \equiv \dot{f}(X_i, \beta)|Y_i - f(X_i, \beta)|. \]

For \( j \in J \), let \( Y_j^* = Y_j - f(X_j, \beta - \beta^0) + f(X_j, \beta) \) and

\[ g(X_j, \beta) \equiv g_j(\beta) \equiv \dot{f}(X_j, \beta)|Y_j^* - f(X_j, \beta)|. \]

We remark that, under the hypothesis \( H_0 \), we have \( g_i(\beta^0) = \dot{f}(X_i, \beta^0)\varepsilon_i \) for \( i \in I \) and \( g_j(\beta^0) = \dot{f}(X_j, \beta^0)\varepsilon_j \) for \( j \in J \). For all \( i = 1, \ldots, n \), we have \( E[g_i(\beta^0)] = 0_d \). We denote by \( \sigma^2_1 \) and \( \sigma^2_2 \) the variance of \( \varepsilon_i \) and \( \varepsilon_j \), respectively. We consider also the \( d \times d \) matrix

\[ V \equiv E[\dot{f}(X_i, \beta^0)\dot{f}^t(X_i, \beta^0)]. \]

In order, to introduce the maximum empirical likelihood method in the following section, let \( y_1, \ldots, y_k, y_{k+1}, \ldots, y_n \) be observations for the random variables \( Y_1, \ldots, Y_k, Y_{k+1}, \ldots, Y_n \). Corresponding to the sets \( I \) and \( J \), the probability vectors \( (p_1, \ldots, p_k) \) and \( (q_{k+1}, \ldots, q_n) \). These vectors contain the probability to observe the value \( y_i \) (respectively \( y_j \)) for the dependent variable \( Y_i \).
respectively \( Y_j \): \( p_i \equiv \mathbb{P}[Y_i = y_i] \), for \( i = 1, \cdots, k \) and \( q_j \equiv \mathbb{P}[Y_j = y_j] \), for \( j = k+1, \cdots, n \). Obviously, these probabilities satisfy the relations \( \sum_{i \in I} p_i = 1 \) and \( \sum_{j \in J} q_j = 1 \).

3. Model with complete data

In this section, we suppose that, for the nonlinear model given by (1), the response variable \( Y_i \) is observed for each \( i = 1, \cdots, n \). We will construct the empirical likelihood ratio statistic and show that this statistic has a \( \chi^2 \) asymptotic distribution, which allows us to construct the confidence regions for \( \gamma \).

3.1. Test statistic

In this subsection, we formulate the empirical likelihood ratio statistic will be used to construct the confidence region for \( \gamma = \beta - \beta_1 \), or for testing hypothesis \( H_0 \) given by (2) against the alternative \( H_1 \) given by (3).

Under hypothesis \( H_0 \), we have \( \gamma_0 = \beta^0 - \beta_1 \). Since \( \beta^0 \) is unknown we use the notation \( \beta \). Then, the profile empirical likelihood for \( \gamma \), evaluated at \( \gamma_0 \) under \( H_0 \) is defined as

\[
\mathcal{R}_{nk}(\gamma_0, \beta) = \sup_{(p_1, \cdots, p_k, q_{k+1}, \cdots, q_n)} \left\{ \prod_{i \in I} p_i \prod_{j \in J} q_j; \sum_{i \in I} p_i = 1, \sum_{j \in J} q_j = 1, \sum_{i \in I} p_i g_i(\beta) = 0_d, \sum_{j \in J} q_j g_j(\beta) = 0_d \right\}.
\]

Let us consider the least squares estimators \( \hat{\beta} \) and \( \hat{\beta}_1 \) of \( \beta \) and \( \beta_1 \), on the observations corresponding to the sets \( I \) and \( J \)

\[
\hat{\beta} = \arg \min_{\beta} \sum_{i \in I} (Y_i - f(X_i, \beta))^2, \quad \hat{\beta}_1 = \arg \min_{\beta_1} \sum_{j \in J} (Y_j - f(X_j, \beta_1))^2.
\]

Then, the corresponding empirical log-likelihood function can be written as

\[
\hat{Z}_{nk}(\gamma_0, \beta) = -2 \log[\mathcal{R}_{nk}(\gamma_0, \beta)/\mathcal{R}_{nk}(\hat{\gamma}, \hat{\beta})]
\]

\[
= \sup_{(p_1, \cdots, p_k, q_{k+1}, \cdots, q_n)} \left\{ \sum_{i \in I} \log(kp_i) + \sum_{j \in J} \log((n-k)q_j); \sum_{i \in I} p_i = 1, \sum_{j \in J} q_j = 1, \sum_{i \in I} p_i g_i(\beta) = 0_d, \sum_{j \in J} q_j g_j(\beta) = 0_d \right\},
\]

where \( \hat{\gamma} = \hat{\beta} - \hat{\beta}_1 \), with \( \hat{\beta}, \hat{\beta}_1 \) given by the equation (4).

In order that the parameters belong a bounded set, in the place of \( k \), we consider \( \theta_{nk} \equiv k/n \). Using the Lagrange multiplier method, we consider the
following random process \( \sum_{i \in I} \log p_i + \eta(\sum_{i \in I} p_i - 1) - (n \theta_{nk}) \lambda_1^i \sum_{i \in I} p_i g_i(\beta) \), with \( \lambda_1 \) and \( \eta \) the Lagrange multipliers, \( \lambda_1 \in \mathbb{R}^d \) and \( \eta \in \mathbb{R} \). Taking derivative with respect to \( p_i \) of this process equal to zero, we obtain

\[ p_i = \frac{1}{n \theta_{nk} \lambda_1^i g_i(\beta) - \eta}. \]

Then, \( 1 + \eta p_i - n \theta_{nk} p_i \lambda_1^i g_i(\beta_1) = 0 \) and we obtain that \( \eta = -n \theta_{nk} \). Hence, the probability \( p_i \) becomes

\[ p_i = \frac{1}{n \theta_{nk}(1 + \lambda_1^i g_i(\beta))}. \]  

(6)

Similarly, for \( j \in J \), we can obtain

\[ q_j = \frac{1}{n(1 - \theta_{nk})(1 - (1 - \theta_{nk})^{-1} \lambda_2^j g_j(\beta))}, \]  

(7)

where \( \lambda_2 \in \mathbb{R}^d \) is the Lagrange multiplier.

Using the equations (6) and (7), the statistic of (5) becomes

\[ \tilde{Z}_{nk}(\gamma_0, \lambda_1, \lambda_2, \beta) = 2 \left( \sum_{i \in I} (1 + \frac{1}{\theta_{nk}} \lambda_1^i g_i(\beta)) + \sum_{j \in J} (1 - \frac{1}{1 - \theta_{nk}} \lambda_2^j g_j(\beta)) \right). \]  

(8)

In order to have single parameters denoted by \( \lambda \), we restrict the study to a particular case, when \( \lambda_1 \) and \( \lambda_2 \) satisfy the constraint \( V_{1n}(\beta) \lambda_1 = V_{2n}(\beta) \lambda_2 \), with

\[ V_{1n}(\beta) = (n \theta_{nk})^{-1} \sum_{i \in I} g_i(\beta), \quad V_{2n}(\beta) = (n(1 - \theta_{nk}))^{-1} \sum_{j \in J} g_j(\beta). \]

In the case of the true parameter \( \beta^0 \), this two last matrices are denoted \( V_{1n}^0 = V_{1n}(\beta^0) \) and \( V_{2n}^0 = V_{2n}(\beta^0) \). Considering constraint \( V_{1n}(\beta) \lambda_1 = V_{2n}(\beta) \lambda_2 \), the statistic (8) becomes

\[ \tilde{Z}_{nk}(\gamma_0, \lambda, \beta) = 2 \left( \sum_{i \in I} (1 + \frac{1}{\theta_{nk}} \lambda_1^i g_i(\beta)) + \sum_{j \in J} (1 - \frac{1}{1 - \theta_{nk}} \lambda_2^j V_{2n}(\beta) g_j(\beta)) \right). \]  

(9)

We will study the maximum of empirical log-likelihood test statistic \( \tilde{Z}_{nk}(\gamma_0, \lambda, \beta) \). Then, we calculate the maximum of empirical log-likelihood test statistic \( \tilde{Z}_{nk}(\gamma_0, \lambda, \beta) \).

\[ \phi_{1n}(\gamma_0, \lambda, \beta) = \frac{\partial \tilde{Z}_{nk}(\lambda, \gamma_0, \beta)}{2 \partial \lambda} \]

\[ = \sum_{i \in I} \frac{g_i(\beta)}{\theta_{nk} + \lambda_1^i g_i(\beta)} - \sum_{j \in J} \frac{V_{1n}(\beta) V_{2n}^{-1}(\beta) g_j(\beta)}{1 - \theta_{nk} - \lambda_1^i V_{1n}(\beta) V_{2n}^{-1}(\beta) g_j(\beta)}. \]  

(10)
\[
\hat{\phi}_{2n}(\gamma_0, \lambda, \beta) = \frac{\partial \hat{Z}_{nk}(\lambda, \gamma_0, \beta)}{\partial \beta} = \sum_{i \in I} \hat{g}_i(\beta)\lambda^t - \sum_{j \in J} \frac{\partial (V_{1n}(\beta)V_{2n}(\beta)g_j(\beta))/\partial \beta}{1 - \theta_{nk} - \lambda^t V_{1n}(\beta)V_{2n}(\beta)g_j(\beta)} \lambda^t.
\]

Then, solving the system \( \hat{\phi}_{1n}(\gamma_0, \lambda, \beta) = 0_d \) and \( \hat{\phi}_{2n}(\gamma_0, \lambda, \beta) = 0_d \), the obtained solutions \( \hat{\lambda}(\theta_{nk}) \) and \( \hat{\beta}(\theta_{nk}) \) are the maximizers of the statistic \( Z \).

We emphasize that, compared with a linear model, in our case, matrix \( V_{1n}(\beta) \), \( V_{2n}(\beta) \) and derivative \( g(\beta) \) depend on \( \beta \). These, besides the nonlinearity of \( g(\beta) \) involve difficulties in the study of the statistic \( \hat{Z}_{nk}(\gamma_0, \lambda, \beta) \) and of the solutions \( \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk}) \).

Remark 1 To acquire the symmetric form of the statistic \( \hat{Z}_{nk}(\gamma_0, \lambda, \beta) \) given by \( f \), which makes the arguments more concise, we consider the following notations :

\[
z_i(\beta) = M_n^{1/2}(\beta) V_{1n}(\beta) \lambda^t, \quad z_j(\beta) = M_n^{1/2}(\beta) V_{2n}(\beta) g_j(\beta),
\]

\[
M_n(\beta) = \theta_{nk}(1 - \theta_{nk}) V_{1n}(\beta) H_n(\beta) V_{2n}(\beta),
\]

\[
H_n(\beta) = \left[ \theta_{nk} \sigma^2 \lambda^t V_{1n}(\beta) + (1 - \theta_{nk}) \sigma^2 \lambda^t V_{2n}(\beta) \right]^{-1}.
\]

Taking into account the above notations, we consider instead of \( \hat{Z}_{nk}(\gamma_0, \lambda, \beta) \) given by \( f \), the following test statistic

\[
Z_{nk}(\gamma_0, \lambda, \beta) = 2 \left[ \sum_{i \in I} \log(1 + \frac{1}{\theta_{nk}} \lambda^t z_i(\beta)) + \sum_{j \in J} \log(1 - \frac{1}{1 - \theta_{nk}} \lambda^t z_j(\beta)) \right].
\]

Based on \( f \), we can derive the following score equations to get the estimators \( (\hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) \) of \( (\lambda, \beta) \):

\[
\begin{align*}
\phi_{1n}(\gamma_0, \lambda, \beta) &\equiv \frac{\partial Z_{nk}(\gamma_0, \lambda, \beta)}{\partial \lambda} = \sum_{i \in I} \frac{z_i(\beta) \lambda^t}{\theta_{nk} + \lambda^t z_i(\beta)} - \sum_{j \in J} \frac{z_j(\beta) \lambda^t}{1 - \theta_{nk} - \lambda^t z_j(\beta)}, \\
\phi_{2n}(\gamma_0, \lambda, \beta) &\equiv \frac{\partial Z_{nk}(\gamma_0, \lambda, \beta)}{\partial \beta} = \sum_{i \in I} \frac{z_i(\beta) \lambda^t}{\theta_{nk} + \lambda^t z_i(\beta)} - \sum_{j \in J} \frac{z_j(\beta) \lambda^t}{1 - \theta_{nk} - \lambda^t z_j(\beta)},
\end{align*}
\]

where \( z_i(\beta) \) and \( z_j(\beta) \) are the derivative with respect to \( \beta \) of \( z_i(\beta) \) and \( z_j(\beta) \) respectively.

In applications, the error variance \( \sigma^2 \) can be estimated by \( (n\theta)^{-1} \sum_{i \in I} [Y_i - f(X_i, \beta)]^2 \), and \( \sigma^2 \) can be estimated by \( (n(1 - \theta))^{-1} \sum_{j \in J} [Y_j - f(X_j, \beta_1)]^2 \), with \( \beta \) and \( \beta_1 \) given by relation \( f \).
3.2. Asymptotic behaviour of the statistic $Z_{nk}$

In this subsection, we will study the asymptotic behaviour of the statistic $Z_{nk}$ given by equation (12) under the null hypothesis $H_0$ given by (2) and we show that $\hat{\lambda}(\theta_{nk})$ and $\hat{\beta}(\theta_{nk})$, the solutions of the score equations $\phi_{1n}(\gamma_0, \lambda, \beta) = 0_d$ and $\phi_{2n}(\gamma_0, \lambda, \beta) = 0_d$ given by the relation (13), have suitable properties.

We require the following assumptions for the next theorem:

(A5) The matrices $V_{1n}(\beta)$ and $V_{2n}(\beta)$ are non singular for any $X_i \in \mathcal{Y}$ and $\beta$ in a neighborhood of $\beta^0$ and their determinants are bounded for sufficiently large $n$.

(A6) $\sup_{\beta} \mathbb{E}[\|f(X, \beta)\|_2^2] < \infty$ and for all $1 \leq j, k \leq d$, $\sup_{\beta} \mathbb{E}[\|\partial^2 f(X, \beta) / \partial \beta_j \partial \beta_k\|] < \infty$, for some $s > 2$.

Assumption (A5) assures that the matrices $V_{1n}(\beta)$ and $V_{2n}(\beta)$ are uniformly nonsingular and bounded for $n$ larger than some integer. Assumption (A6) is a necessary moment condition for statistical inference and it is also employed in paper of Boldea and Hall [2].

By the next proposition, we show that $\hat{\lambda}(\theta_{nk})$ and $\hat{\beta}(\theta_{nk})$, the solutions of the score equations $\phi_{1n}(\gamma_0, \lambda, \beta) = 0_d$ and $\phi_{2n}(\gamma_0, \lambda, \beta) = 0_d$ given by the relation (13), have suitable properties. More precisely, we show that $\|\hat{\lambda}(\theta_{nk})\|_2 \to 0$, as $n \to \infty$ and that $\hat{\beta}(\theta_{nk})$ is a consistent estimator of $\beta^0$, under hypothesis $H_0$.

Proposition 1 Under the null hypothesis $H_0$, if the assumptions (A1)-(A6) are satisfied, then we have $\hat{\lambda}(\theta_{nk}) = O_P(n^{-1/2})$ and $\hat{\beta}(\theta_{nk}) = o_P(1)$, where $(\hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk}))$ is the solution of the system (13).

The proof of Proposition 1 is given in Appendix. It is similar to the Theorem 2 of [8], but we have an important modifications due to that, in this paper the test statistic $Z_{nk}(\gamma_0, \lambda, \beta)$ depend to $z_i(\beta)$ (respectively $z_j(\beta)$) and not $g_i(\beta)$ (respectively $g_j(\beta)$).

The following result is a generalization of the nonparametric version of Wilk’s theorem for the empirical likelihood ratio defined by (12).

Theorem 1 Suppose that assumptions (A1)-(A6) hold. Under the null hypothesis $H_0$, the statistic $Z_{nk}(\gamma_0, \lambda, \beta)$ given by (12) converge, as $n \to \infty$, to a chi-squared distribution with $d$ degrees of freedom, where $d$ is the dimension of $\beta$.

From Theorem 1 we can construct confidence regions for $\gamma$ as follow:

$$CR_{\alpha} = \{\gamma \in \mathbb{R}^d : Z_{nk}(\gamma_0, \lambda, \beta) < c_{1-\alpha, d}\},$$
where \( c_{1-\alpha,d} \) is the \((1-\alpha)\) quantile of the chi-squared distribution with \( d \) degrees of freedom.

4. Model with missing response data

In this section, for the model (1), we suppose that some values of \( Y \) may be missing and \( X_i \) is observed completely. That is, we obtain an incomplete sample \( \{(X_i, Y_i, \delta_i)_{1 \leq i \leq n}\} \) from model (1), where all the \( X_i \) is observed, \( (\delta_i)_{1 \leq i \leq n} \) is a sequence of random variables with \( \delta_i = 0 \) if \( Y_i \) is missing and \( \delta_i = 1 \) otherwise. We assume that \( Y_i \) is missing at random (MAR). That is, 
\[
P[\delta_i = 1|X_i, Y_i] = P[\delta_i = 1|X_i], \quad \text{for all } 1 \leq l \leq n.
\]

We consider the selective probabilities functions defined as 
\[
\pi_1(x_i) = P[\delta_i = 1|X_i = x_i], \quad \text{for } i \in I \\
\pi_2(x_j) = P[\delta_j = 1|X_j = x_j], \quad \text{for } j \in J.
\]

We consider the supposition \( \pi_1(X_i), \pi_2(X_j) > 0 \) which is a common supposition in the literature, for example see the papers of Sun et al. [23] and Xue [25].

The literature on statistical analysis of data with missing values has flourished since the early 1970, spurred by advances in computer-technology that made previously laborious numerical calculations a simple matter. In practice, however, response variables are usually missing 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. Actually, missingness of responses is very common in opinion polls, market research surveys and many scientific experiments. When some responses are missing, the existing methods in the literature are not applicable any more.

A nonlinear model based on missing at random (MAR), has been considered by various authors. Muller [16] constructed an efficient estimator for expectation \( IE[h(X, Y)] \) using a efficient estimator of parameters, with \( h \) is a known square integrable function. The mean response \( E(Y) \) is a special case. Ciperca [7] constructed the empirical likelihood ratios using complete-case and imputed values. The basic idea in imputation is to "fill in" missing \( Y \) values with "appropriate" values to create a completed data set, thereby allowing standard methods to be applied. However, the imputed data are not i.i.d. because a plug-in estimator is used. The empirical log-likelihood ratio under imputation is asymptotically distributed as a scaled chi-square variable.

For the rest of this section, an empirical likelihood method is used to study model (1) under missing response data. We are interested to construct the confidence region for \( \gamma = \beta - \beta_1 \), based on the data \( (X_i, Y_i, \delta_i)_{1 \leq i \leq n} \), or testing the hypothesis
\[
H_0 : \gamma = \gamma_0.
\]
A class of empirical log-likelihood ratio functions for \( \gamma \) are defined that include the following three types: a profile empirical likelihood ratio for \( \gamma \) with
complete-case data, a weighted empirical likelihood ratio for $\gamma$, an empirical likelihood ratio for $\gamma$ based on imputed values.

In the missing response data case, the least squares estimators $\hat{\beta}$ and $\hat{\beta}_1$ of $\beta$ and $\beta_1$, are respectively

$$\hat{\beta} = \arg \min_{\beta} \sum_{i \in I} \delta_i (Y_i - f(X_i, \beta))^2, \quad \hat{\beta}_1 = \arg \min_{\beta_1} \sum_{j \in J} \delta_j (Y_j - f(X_j, \beta_1))^2.$$  \hspace{1cm} (14)

We recall that, $\sigma_1^2$ and $\sigma_2^2$ denote the variance of $\varepsilon_i$ and $\varepsilon_j$, respectively. In the missing response data case, $\sigma_1^2$ and $\sigma_2^2$ are estimated respectively by $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$, where

$$\tilde{\sigma}_1^2 = \frac{\sum_{i \in I} \delta_i (Y_i - f(X_i, \hat{\beta}))^2}{\sum_{i \in I} \delta_i}, \quad \tilde{\sigma}_2^2 = \frac{\sum_{j \in J} \delta_j (Y_j - f(X_j, \hat{\beta}_1))^2}{\sum_{j \in J} \delta_j},$$ \hspace{1cm} (15)

with $\hat{\beta}$ and $\hat{\beta}_1$ are given by the relation (14).

4.1. Test statistics

Firstly, we give the empirical likelihood based on complete-case data, i.e, excluding missing data. In the regression context, this usually means complete-case analysis : excluding all units for which the outcome or any of the inputs are missing.

Two problems arise with complete-case analysis. First, if the units with missing values differ systematically from the completely observed cases, this could bias the complete-case analysis. Second, if many variables are included in a model, there may be very few complete cases, so that most of the data would be discarded for the sake of a simple analysis.

For $j \in J$, let $\bar{Y}_j = Y_j - f(X_j, \beta - \gamma^0) + f(X_j, \beta)$, if $Y_j$ non missing and $\bar{Y}_j$ equal to any finite value if $Y_j$ missing. We define the two following $d$–random vectors

$$\left\{ \begin{array}{l} g_{i,C}(\beta) = \delta_i f_i(\beta)(Y_i - f(X_i, \beta)) \quad i \in I, \\
 g_{j,C}(\beta) = \delta_j f_j(\beta)(\bar{Y}_j - f_j(\beta)) \quad j \in J. \end{array} \right.$$  \hspace{1cm} (16)

Let us consider in this case, the following matrices

$$V_{1n,C}(\beta) = \frac{1}{n\theta_{nk}} \sum_{i \in I} g_{i,C}(\beta), \quad V_{2n,C}(\beta) = \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} g_{j,C}(\beta).$$

Similarly as in Section 2, and by the same argument given in Remark 1, for $\theta_{nk} = k/n$, the corresponding empirical likelihood ratio statistic, for constructing confidence region or for testing hypothesis in the complete-case data method, is

$$Z_{nk,C}(\gamma_0, \lambda C, \beta) = 2 \sum_{i \in I} \log(1 + \frac{1}{\theta_{nk}} \lambda_i \lambda_i^T z_{i,C}(\beta)) + \sum_{j \in J} \log(1 - \frac{1}{1 - \theta_{nk}} \lambda_j \lambda_j^T z_{j,C}(\beta)), \hspace{1cm} (17)$$
with,
\[ z_{i,C}(\beta) = M_{n,C}^{-1}(\beta) V_{1n,C}(\beta) g_{i,C}(\beta), \quad z_{j,C}(\beta) = M_{n,C}^{-1}(\beta) V_{2n,C}(\beta) g_{j,C}(\beta), \]
and
\[ M_{n,C}(\beta) = \theta_{nk}(1 - \theta_{nk}) V_{1n,C}(\beta) H_{n,C}(\beta) V_{2n,C}(\beta), \]
and
\[ H_{n,C}(\beta) = \left[ \frac{\sigma_{1}^{2}(1 - \theta_{nk})}{n \theta_{nk}} V_{2n,C}(\beta) V_{1n,C}(\beta) \sum_{i \in I} \delta_{i} \pi_{1}(X_{i}) \left[ \hat{f}_{i}(\beta) (Y_{i} - f_{i}(\beta)) - \hat{f}_{i}(\beta) \hat{f}_{i}^{t}(\beta) \right] \right. 
\[ + \frac{\sigma_{2}^{2} \theta_{nk}}{n(1 - \theta_{nk})} \sum_{j \in J} \delta_{j} \pi_{2}(X_{j}) \left[ \hat{f}_{j}(\beta) (\hat{Y}_{j} - f_{j}(\beta)) - \hat{f}_{j}(\beta) \hat{f}_{j}^{t}(\beta) \right] V_{1n,C}(\beta) \right]^{-1}. \]

The variances \( \sigma_{1}^{2} \) and \( \sigma_{2}^{2} \) can be estimated respectively by \( \hat{\sigma}_{1}^{2} \) and \( \hat{\sigma}_{2}^{2} \), given by [15]. In this case, the score functions of test statistic (17) are

\[
\begin{cases}
\phi_{1n,C}(\gamma_{0}, \lambda_{C}, \beta) = \sum_{i \in I} \frac{z_{i,C}(\beta)}{\theta_{nk} + \lambda_{C} z_{i,C}(\beta)} - \sum_{j \in J} \frac{z_{j,C}(\beta)}{1 - \theta_{nk} - \lambda_{C} z_{j,C}(\beta)}, \\
\phi_{2n,C}(\gamma_{0}, \lambda_{C}, \beta) = \sum_{i \in I} \frac{z_{i,C}(\beta)}{\theta_{nk} + \lambda_{C} z_{i,C}(\beta)} - \sum_{j \in J} \frac{z_{j,C}(\beta)}{1 - \theta_{nk} - \lambda_{C} z_{j,C}(\beta)},
\end{cases}
\tag{18}
\]
where \( \hat{z}_{i,C}(\beta) \) and \( \hat{z}_{j,C}(\beta) \) are the derivative with respect to \( \beta \) of \( z_{i,C}(\beta) \) and \( z_{j,C}(\beta) \) respectively. Then, solving the system \( \phi_{1n,C}(\gamma_{0}, \lambda_{C}, \beta) = 0_{d} \) and \( \phi_{2n,C}(\gamma_{0}, \lambda_{C}, \beta) = 0_{d} \) given by (18), we obtain \( \hat{\lambda}_{C}(\theta_{nk}) \) and \( \hat{\beta}_{C}(\theta_{nk}) \) the maximizers of the statistic (17).

The selective probabilities \( \pi_{1}(X_{i}) \) and \( \pi_{2}(X_{j}) \) are considered as known. If they are unknown, we consider the nonlinear estimators \( \hat{\pi}_{1}(X_{i}) \) and \( \hat{\pi}_{2}(X_{j}) \) for \( \pi_{1}(X_{i}) \) and \( \pi_{2}(X_{j}) \), respectively, given by

\[
\begin{align*}
\hat{\pi}_{1}(X_{i}) &= \frac{\sum_{j \in J} \delta_{j} K_{1}(X_{i} - X_{j}/h_{1n})}{\max \{ \sum_{j \in J} K_{1}(X_{i} - X_{j}/h_{1n}) \}}, \quad i \in I, \\
\hat{\pi}_{2}(X_{j}) &= \frac{\sum_{i \in I} \delta_{i} K_{2}(X_{j} - X_{i}/h_{2n})}{\max \{ \sum_{i \in I} K_{2}(X_{j} - X_{i}/h_{2n}) \}}, \quad j \in J.
\end{align*}
\tag{19}
\]

Here, \( h_{1n} \) and \( h_{2n} \) are positive sequences tending towards 0 as \( n \to \infty \). \( K_{1} \) and \( K_{2} \) are kernel functions defined in \( \mathbb{R}^{d} \).
The bandwidths \( h_{1n} \) and \( h_{2n} \) satisfies the following :

\[ n \theta_{nk} h_{1n}^{4 \max \{2,d-1\}} \to 0 \text{ and } (n - n \theta_{nk}) h_{2n}^{4 \max \{2,d-1\}} \to 0, \text{ as } n \to \infty. \]

The kernel functions \( K_{1} \) and \( K_{2} \) satisfy the classical condition :

\[ (A7) \text{ There exist positive constants } C_{1}, C_{2}, C_{3}, C_{4}, \rho_{1} \text{ and } \rho_{2}, \text{ such that, for any vector } v, \quad C_{1} \mathbb{1}_{\|v\| \leq \rho_{1}} \leq K_{1}(v) \leq C_{2} \mathbb{1}_{\|v\| \leq \rho_{2}}, \text{ and } C_{3} \mathbb{1}_{\|v\| \leq \rho_{1}} \leq K_{2}(v) \leq C_{4} \mathbb{1}_{\|v\| \leq \rho_{2}}. \]
Condition (A8) is also imposed in the papers of Wei et al. [24] and Xue [25]. Concerning the selective probabilities functions \( \pi_1(X_i) \) and \( \pi_2(X_j) \), let us consider the following regularity hypotheses:

(A9) \( \pi_1(X_i) \) and \( \pi_2(X_j) \) have bounded partial derivatives up to order \( \max(2, d-1) \) almost everywhere.

Conditions (A7)-(A9) are the usual assumptions for the convergence rate of the kernel estimation method, for example, see the paper of Wei et al. [24].

Now, we give the weighted empirical likelihood method. As discussed previously, complete-case analysis can yield biased estimates because the sample of observations that have no missing data might not be representative of the full sample. We could build a model to predict the nonresponse in that variable using all the other variables. The inverse of predicted probabilities of response from this model could then be used as survey weights to make the complete-case sample representative (along the dimensions measured by the other predictors) of the full sample. This method becomes more complicated when there is more than one variable with missing data.

In order to obtain the weighted empirical likelihood statistic, we use the inverse probability weighted approach for missing data analysis, which was used by Horvitz and Thompson [9] for missing data analysis.

We define the two following \( d \)-random vectors

\[
\begin{align*}
\{ g_{i,W}(\beta) = \frac{d}{\pi_1(X_i)} f_i(\beta)(Y_i - f_i(\beta)) & \quad i \in I, \\
g_{j,W}(\beta) = \frac{d}{\pi_2(X_j)} f_j(\beta)(Y_j - f_j(\beta)) & \quad j \in J.
\end{align*}
\]

We recall that, \( \hat{Y}_j = Y_j - f(X_j, \beta - \gamma^0) + f(X_j, \beta) \) for \( j \in J \), if \( Y_j \) non missing and \( \hat{Y}_j \) equal to any finite value if \( Y_j \) missing.

Let us consider in this case, the following matrices

\[
\begin{align*}
V_{1n,W}(\beta) & = \frac{1}{n\theta nk} \sum_{i \in I} g_{i,W}(\beta), & V_{2n,W}(\beta) & = \frac{1}{n(1 - \theta nk)} \sum_{j \in J} g_{j,W}(\beta).
\end{align*}
\]

Like as in the complete-case data, using in this case the similar argument of Remark 1 the test statistic for the weighted method is

\[
Z_{nk,W}(\gamma_0, \lambda_W, \beta) = 2 \sum_{i \in I} \log(1 + \frac{1}{\theta nk} \lambda_W g_{i,W}(\beta)) + \sum_{j \in J} \log(1 - \frac{1}{1 - \theta nk} \lambda_W z_{j,W}(\beta)),
\]

where

\[
\begin{align*}
z_{i,W}(\beta) & = M_{n,W}(\beta) V_{1n,W}(\beta) g_{i,W}(\beta), \quad z_{j,W}(\beta) = M_{n,W}(\beta) V_{2n,W}(\beta) g_{j,W}(\beta),
\end{align*}
\]

\[C_1 \|
u\| \leq \rho_2.\]
\[ M_{n,W}(\beta) = \theta_{nk}(1 - \theta_{nk})V_{1n,W}(\beta)H_{n,W}(\beta)V_{2n,W}(\beta), \]

and

\[ H_{n,W}(\beta) = \left[ \frac{\sigma_1^2(1 - \theta_{nk})}{n\theta_{nk}}V_{2n,W}(\beta) + \frac{\sigma_2^2\theta_{nk}}{n(1 - \theta_{nk})}V_{1n,W}(\beta) \right]^{-1}. \]

The variances \( \sigma_1^2 \) and \( \sigma_2^2 \) can be estimated respectively by \( \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \), given by [15].

The score function of test statistic (21) is given by (15).

Then, solving the system \( \phi_{1n,W}(\gamma_0, \lambda_W, \beta) = 0 \) and \( \phi_{2n,W}(\gamma_0, \lambda_W, \beta) = 0 \) given by (22), we obtain \( \hat{\lambda}_W(\theta_{nk}) \) and \( \hat{\beta}_W(\theta_{nk}) \) the maximizers of the statistic (21).

Finally, we give the empirical likelihood with imputed values method. For the profile empirical likelihood with complete-case data and the weighted empirical likelihood, the information contained in the data is not explored fully. Since incomplete-case data are discarded in constructing the empirical likelihood ratio, the coverage accuracies of confidence regions are reduced when there are plenty of missing values. To resolve the issue, we use nonlinear regression imputation to impute \( Y_i \) (or \( Y_j \)) if \( Y_i \) (or \( Y_j \)) is missing. We introduce the forecast of \( Y_i \), for \( i = 1, \ldots, n \), constructed using the least square estimators for the parameters \( \beta \) and \( \beta_1 \) and a nonparametric estimators for probabilities \( \pi_1(X_i) \) and \( \pi_2(X_j) \).

\[
Y_{i,m} = \frac{\delta_i}{\pi_1(X_i)}Y_i + \left( 1 - \frac{\delta_i}{\pi_1(X_i)} \right)f(X_i, \hat{\beta}), \quad i \in I,
\]

\[
Y_{j,m} = \frac{\delta_j}{\pi_2(X_j)}Y_j + \left( 1 - \frac{\delta_j}{\pi_2(X_j)} \right)f(X_j, \hat{\beta}_1), \quad j \in J,
\]

where, \( \hat{\beta} \) and \( \hat{\beta}_1 \) given by the relation (14).

For \( j \in J \), let us consider \( Y_{j,m}^* = Y_j - f(X_j, \beta - \gamma^0) + f(X_j, \beta) \). The auxiliary random vectors are defined by

\[
\begin{cases}
\mathbf{g}_i,m(\beta) = \hat{f}_i(\beta)(Y_{i,m} - f_i(\beta)) & i \in I, \\
\mathbf{g}_j,m(\beta) = \hat{f}_j(\beta)(Y_{j,m}^* - f_j(\beta)) & j \in J.
\end{cases}
\]
Let also the following matrices
\[
V_{1n,Im}(\beta) = \frac{1}{n\theta_{nk}} \sum_{i \in I} \dot{g}_{i,Im}(\beta), \quad V_{2n,Im}(\beta) = \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} \dot{g}_{j,Im}(\beta).
\]

Like as in the two above methods, the test statistic for the imputed method is
\[
Z_{nk,Im}(\gamma_0, \lambda_{Im}, \beta) = 2 \left[ \sum_{i \in I} \log \left( 1 + \frac{1}{\theta_{nk}} \lambda_{1m}^t z_{i,Im}(\beta) \right) + \sum_{j \in J} \log \left( 1 - \frac{1}{1 - \theta_{nk}} \lambda_{Im}^t z_{j,Im}(\beta) \right) \right],
\]

where,
\[
z_{i,Im}(\beta) = M_{n,Im}(\beta)^{-1} \sum_{i \in I} \frac{1}{\pi_1(X_i)} \left[ f_i(\beta)(Y_i,Im - f_i(\beta)) - \dot{f}_i(\beta) f_i(\beta) \right] V_{1n,Im}(\beta) V_{1n,Im}^t(\beta) \cdot \left[ f_i(\beta)(Y_i,Im - f_i(\beta)) - \dot{f}_i(\beta) f_i(\beta) \right] V_{1n,Im}^{-1}(\beta),
\]

and
\[
H_{nk,Im}(\beta) = \frac{\sigma_1^2}{n\theta_{nk}} \sum_{i \in I} \frac{1}{\pi_1(X_i)} \left[ f_i(\beta)(Y_i,Im - f_i(\beta)) - \dot{f}_i(\beta) f_i(\beta) \right] \cdot \left[ f_i(\beta)(Y_i,Im - f_i(\beta)) - \dot{f}_i(\beta) f_i(\beta) \right] V_{1n,Im}^{-1}(\beta) V_{1n,Im}^t(\beta) \cdot \left[ f_i(\beta)(Y_i,Im - f_i(\beta)) - \dot{f}_i(\beta) f_i(\beta) \right] V_{1n,Im}^{-1}(\beta).
\]

The variances \(\sigma_1^2\) and \(\sigma_2^2\) can be still be estimated by \ref{15}.

The score functions to test statistic of \ref{24} is
\[
\left\{
\begin{array}{l}
\phi_{1n,Im}(\gamma_0, \lambda_{Im}, \beta) = \sum_{i \in I} \frac{z_{i,Im}(\beta)}{\theta_{nk} + \lambda_{1m}^t z_{i,Im}(\beta)} - \sum_{j \in J} \frac{z_{j,Im}(\beta)}{1 - \theta_{nk} - \lambda_{Im}^t z_{j,Im}(\beta)}, \\
\phi_{2n,Im}(\gamma_0, \lambda_{Im}, \beta) = \sum_{i \in I} \frac{z_{i,Im}(\beta)}{\theta_{nk} + \lambda_{1m}^t z_{i,Im}(\beta)} - \sum_{j \in J} \frac{z_{j,Im}(\beta)}{1 - \theta_{nk} - \lambda_{Im}^t z_{j,Im}(\beta)},
\end{array}
\right.
\]

(25)

where \(z_{i,Im}(\beta)\) and \(z_{j,Im}(\beta)\) are the derivative with respect to \(\beta\) of \(z_{i,Im}(\beta)\) and \(z_{j,Im}(\beta)\) respectively.

Then, solving the system \(\phi_{1n,Im}(\gamma_0, \lambda_{Im}, \beta) = 0_d\) and \(\phi_{2n,Im}(\gamma_0, \lambda_{Im}, \beta) = 0_d\) given by \ref{25}, we obtain \(\lambda_{Im}(\theta_{nk})\) and \(\beta_{Im}(\theta_{nk})\) the maximizers of the statistic \ref{24}.

4.2. Asymptotic behaviours of \(Z_{nk,C}, Z_{nk,W}\) and \(Z_{nk,Im}\)

In this subsection, we study the asymptotic distributions of the empirical likelihood ratios \(Z_{nk,C}, Z_{nk,W}\) and \(Z_{nk,Im}\), given by \ref{17}, \ref{21} and \ref{24} respectively. The main result is given by Theorem \ref{2} where we show that under the null hypothesis \(H_0\), all three statistics have, asymptotically, chi squared distributions.
We require the equivalent to the assumption (A5) given in the no missing response data case:

(A10) The matrices \( V_{1n,C}, V_{2n,C}, V_{1n,W}, V_{2n,W}, V_{1n,Im} \) and \( V_{2n,Im} \) are non singular for any \( X_i \in \mathcal{T} \) and \( \beta \) in a neighborhood of \( \beta^0 \) and their determinants are bounded for sufficiently large \( n \).

Assumption (A10) assures that the matrices \( V_{1n,C}, V_{2n,C}, V_{1n,W}, V_{2n,W}, V_{1n,Im} \) and \( V_{2n,Im} \), are uniformly nonsingular and bounded for \( n \) larger than some integer.

By the next proposition, we show that, \( \hat{\lambda}_C(\theta_{nk}) \) and \( \hat{\beta}_C(\theta_{nk}) \), the solutions of the score equations \( \phi_{1n,C}(\gamma_0, \lambda_C, \beta) = 0 \) and \( \phi_{2n,C}(\gamma_0, \lambda_C, \beta) = 0 \) given by the relation (18), \( \hat{\lambda}_W(\theta_{nk}) \) and \( \hat{\beta}_W(\theta_{nk}) \), the solutions of the score equations \( \phi_{1n}(\gamma_0, \lambda_W, \beta) = 0 \) and \( \phi_{2n}(\gamma_0, \lambda_W, \beta) = 0 \) given by the relation (22) and \( \hat{\lambda}_{Im}(\theta_{nk}) \) and \( \hat{\beta}_{Im}(\theta_{nk}) \), the solutions of the score equations \( \phi_{1n,Im}(\gamma_0, \lambda_{Im}, \beta) = 0 \) and \( \phi_{2n,Im}(\gamma_0, \lambda_{Im}, \beta) = 0 \) given by the relation (25) has suitable properties. More precisely, we show that \( \|\hat{\lambda}_C(\theta_{nk})\|_2 \to 0 \), \( \|\hat{\lambda}_W(\theta_{nk})\|_2 \to 0 \) and \( \|\hat{\lambda}_{Im}(\theta_{nk})\|_2 \to 0 \), as \( n \to \infty \) and that \( \hat{\beta}_C(\theta_{nk}), \hat{\beta}_W(\theta_{nk}) \) and \( \hat{\beta}_{Im}(\theta_{nk}) \) are a consistent estimators of \( \beta^0 \), under hypothesis \( H_0 \).

**Proposition 2** Under the null hypothesis \( H_0 \), if the assumptions (A1)-(A4), (A6)-(A10) are satisfied, for the estimators \( \hat{\lambda}_C(\theta_{nk}), \hat{\beta}_C(\theta_{nk}), \hat{\lambda}_W(\theta_{nk}), \hat{\beta}_W(\theta_{nk}) \) and \( \hat{\lambda}_{Im}(\theta_{nk}), \hat{\beta}_{Im}(\theta_{nk}) \) given by solving the systems (18), (22) and (25) respectively, we have

\[
\hat{\lambda}_C(\theta_{nk}) = \mathcal{O}_p(n^{-1/2}), \quad \hat{\lambda}_W(\theta_{nk}) = \mathcal{O}_p(n^{-1/2}), \quad \hat{\lambda}_{Im}(\theta_{nk}) = \mathcal{O}_p(n^{-1/2}),
\]

\[
\hat{\beta}_C(\theta_{nk}) - \beta^0 = \mathcal{O}_p(1), \quad \hat{\beta}_W(\theta_{nk}) - \beta^0 = \mathcal{O}_p(1), \quad \hat{\beta}_{Im}(\theta_{nk}) - \beta^0 = \mathcal{O}_p(1).
\]

The following theorem gives the asymptotic distribution of the empirical likelihood statistics given by (17), (21) and (24).

**Theorem 2** Suppose that assumptions (A1)-(A4), (A6)-(A10) hold. Under the null hypothesis \( H_0 \), the statistics \( Z_{nk,C}(\gamma_0, \lambda_C, \beta), Z_{nk,W}(\gamma_0, \lambda_W, \beta) \) and \( Z_{nk,Im}(\gamma_0, \lambda_{Im}, \beta) \) all have an asymptotic \( \chi^2_d \) distribution, where \( d \) is the dimension of \( \beta \).

Thus, an asymptotic \((1 - \alpha)\) confidence regions for \( \gamma_0 \), based on the empirical likelihood statistic for the three proposed methods are respectively,

\[
CR_{\alpha,C} = \{ \gamma \in \mathbb{R}^d : Z_{nk,C}(\gamma_0, \lambda_C, \beta) < c_{1-\alpha,d} \},
\]

\[
CR_{\alpha,W} = \{ \gamma \in \mathbb{R}^d : Z_{nk,W}(\gamma_0, \lambda_W, \beta) < c_{1-\alpha,d} \},
\]

\[
CR_{\alpha,Im} = \{ \gamma \in \mathbb{R}^d : Z_{nk,Im}(\gamma_0, \lambda_{Im}, \beta) < c_{1-\alpha,d} \},
\]

where \( c_{1-\alpha,d} \) is the \((1 - \alpha)\) quantile of the chi-squared distribution with \( d \) degrees of freedom.
5. Simulation study

Using a simulation study by Monte Carlo, we now give a results to evaluate the performance of the proposed empirical likelihood confidence regions. Firstly, when the nonlinear regression model with complete-data, secondly, when we have a nonlinear regression model with missing response data. We use the R language for all simulations. The program codes are available from the author.

We consider the following nonlinear function

\[ f(x, \beta) = a \frac{1 - x^b}{b}, \]  

with \( \beta = (a, b) \in [-100, 100] \times [0.1, 20] \), \( X_i = i/1000 \), \( n = 1000 \), \( a_1^0 = 10 \) and \( b_1^0 = 2 \).

We take a three different laws for the error terms,

Case (a) : \( \varepsilon_i = N(0, 1) \) and \( \varepsilon_j = N(0, 1) \),
Case (b) : \( \varepsilon_i = 1/\sqrt{6}(\chi^2(3) - 3) \) and \( \varepsilon_j = 2/\sqrt{6}t(6) \),
Case (c) : \( \varepsilon_i = 2\exp(2) - 1 \) and \( \varepsilon_j = N(0, 1) \),

where \( N(0, 1) \), \( \exp(2) \), \( \chi^2(3) \) and \( t(6) \) are standard normal distribution, exponential distribution with mean 1/2, chi-square distribution with degree of freedom 3 and Student distribution with degree of freedom 6, respectively. The nominal coverage level is \( 1 - \alpha = 0.95 \).

5.1. Model with complete data

For nominal confidence level \( 1 - \alpha = 0.95 \), for 1000 Monte Carlo replications, the Table 1 presents the coverage probabilities (CP) and lengths of the confidence regions (LCR) given by the empirical log-likelihood method on the no-missing case data (Theorem 1):

\[ CR_\alpha = \{ \gamma \in \mathbb{R}^d : Z_{nk}(\gamma_0, \lambda, \beta) < c_{1-\alpha;d} \}, \]

where \( c_{1-\alpha;d} \) is the \( (1 - \alpha) \) quantile of the standard chi-squared distribution with \( d \) degrees of freedom and \( Z_{nk}(\gamma_0, \lambda, \beta) \) is given by equation (12).

In order to calculate the coverage probability (CP), we consider a model under hypothesis \( H_0 \) given by (2) and we count the number of times, on the Monte Carlo replications, when the statistic value does not exceeds the critical value \( c_{1-\alpha;d} \). The lengths of the confidence regions (LCR), for each simulation, designate the difference between value that we are confident of with upper or lower endpoint obtained for the statistic \( Z_{nk} \). We can see that all the coverage probabilities (CP) are very close to 0.95, which indicate the performance of the proposed empirical likelihood confidence region.
Table 1: Coverage probability of confidence regions with three error patterns.

| $k_0$ | Errors $\varepsilon$ | Case(a) | Case(b) | Case(c) |
|-------|------------------------|---------|---------|---------|
|       |                        |         |         |         |
| 300   | CP                     | 0.942   | 0.933   | 0.949   |
|       | LCR                    | 4.876   | 5.682   | 5.470   |
| 500   | CP                     | 0.966   | 0.900   | 0.960   |
|       | LCR                    | 5.151   | 5.233   | 4.939   |
| 700   | CP                     | 0.957   | 0.951   | 0.966   |
|       | LCR                    | 5.757   | 5.531   | 5.794   |

5.2. Model with missing response data

In this subsection, we suppose that the response variable $Y_i$ can be missing at random.

The two next figures represent the first 200 observations of $Y_i$, 'triangle' represents the missing values of $Y_i$ and 'black point' represents the observed values of $Y_i$.

Figure 1: Response variable $Y_i \sim \mathcal{N}(0, 1)$ and $\varepsilon_i \sim \mathcal{E}xp(2)$, n=200.

Throughout this subsection, the Kernel functions are taken as the Epanechnikov Kernel, $K_1(X_i) = 0.75(1 - X_i^2)1_{|X_i| \leq 1}$ and $K_2(X_i) = 0.75(1 - X_i^2)1_{|X_i| \leq 1}$ and the bandwidths $h_{1n} = (n\theta)^{-1/7}$ and $h_{2n} = (n - n\theta)^{-1/7}$ satisfy the condition (A7).

We consider the three following studies of response probabilities under the MAR assumption:

Study 1: $\pi_1(X_i) = 0.8 + 0.2|X_i - 1|$ if $|X_i - 1| \leq 1$ and 0.95 otherwise,
\[ \pi_2(X_j) = 0.8 + 0.2|X_j - 1| \text{ if } |X_j - 1| \leq 1 \text{ and } 0.95 \text{ otherwise.} \]

**Study 2**: \( \pi_1(X_i) = 0.8 \) for all \( X_i \) and \( \pi_2(X_j) = 0.8 \) for all \( X_j \).

**Study 3**: \( \pi_1(X_i) = 0.8 + 0.2|X_i - 1| \) if \( |X_i - 1| \leq 1 \) and 0.95 otherwise, \( \pi_2(X_j) = 0.8 \) for all \( X_j \).

For the studies 1, 2 and 3, the Tables 2, 3, and 4, present respectively, the simulated coverage probabilities (CP) and the interval lengths (LCR) for confidence regions given by the empirical likelihood on the missing response data case (Theorem 2), using the statistics \( Z_{nk,C} \), \( Z_{nk,W} \) and \( Z_{nk,Im} \) given by equations (17), (21) and (24), respectively, with a nominal confidence level \( 1 - \alpha = 0.95 \) and \( n = 1000 \). For each study, we consider the three laws for the errors (a), (b) and (c) given above. We run 1000 replications for each simulation. In all studies and cases, we can see that the coverage probabilities (CP) has a values very close to 0.95, which indicate the performance of the proposed empirical likelihood confidence regions in the three methods.

Table 2: Coverage probability (CP) and interval lengths (LCR) of the three methods for the Study 1, \( n = 1000 \).

| Y          | Errors \( \varepsilon \) | Case(a) | Case(b) | Case(c) |
|------------|--------------------------|---------|---------|---------|
| Comple-Case| CP                        | 0.920   | 0.918   | 0.913   |
|            | LCR                       | 4.450   | 5.249   | 4.988   |
| Weighted   | CP                        | 0.917   | 0.910   | 0.903   |
|            | LCR                       | 5.896   | 5.792   | 4.725   |
| Imputed    | CP                        | 0.921   | 0.926   | 0.915   |
|            | LCR                       | 5.630   | 5.759   | 4.778   |
Table 3: Coverage probability (CP) and interval lengths (LCR) of the three methods for the Study 2, $n = 1000$.

| Y Errors $\varepsilon$ | Case(a) | Case(b) | Case(c) |
|-------------------------|---------|---------|---------|
| Comple-Case CP          | 0.943   | 0.911   | 0.932   |
| LCR                     | 5.963   | 5.837   | 5.869   |
| Weighted CP             | 0.901   | 0.905   | 0.918   |
| LCR                     | 5.981   | 4.717   | 4.778   |
| Imputed CP              | 0.934   | 0.917   | 0.919   |
| LCR                     | 5.566   | 5.809   | 4.594   |

Table 4: Coverage probability (CP) and interval lengths (LCR) of the three methods for the Study 3, $n = 1000$.

| Y Errors $\varepsilon$ | Case(a) | Case(b) | Case(c) |
|-------------------------|---------|---------|---------|
| Comple-Case CP          | 0.925   | 0.943   | 0.914   |
| LCR                     | 5.676   | 5.342   | 5.863   |
| Weighted CP             | 0.902   | 0.924   | 0.913   |
| LCR                     | 4.854   | 5.600   | 5.376   |
| Imputed CP              | 0.922   | 0.928   | 0.939   |
| LCR                     | 5.912   | 5.559   | 5.552   |

6. Appendix

The following lemma will be used in the proof of propositions, theorems and of other lemmas.

**Lemma 1** Let $X = (X_1, \cdots, X_p)$ a random vector (column), with the random variables $X_1, \cdots, X_p$ not necessarily independent and $M = (m_{ij})_{1 \leq i, j \leq p}$ such that $M = XX^t$. If for $j=1, \ldots, p$, we have

\[
P\left[|X_j| \geq \delta_j \right] \leq \eta_j,
\]

then

(i) $P\left[\|X\|_1 \geq p \max_{1 \leq j \leq p} \delta_j \leq \max_{1 \leq j \leq p} \eta_j\right]$, (ii) $P\left[\|X\|_2 \geq \sqrt{p} \max_{1 \leq j \leq p} \delta_j \leq \max_{1 \leq j \leq p} \eta_j\right]$, (iii) $E\left[\|M\|_1 \geq p \max_{1 \leq i,j \leq p} \{\delta_i^2, \delta_j^2\} \right] \leq \max_{1 \leq i,j \leq p} \{\eta_i^2, \eta_j^2\}$,

where $\|M\|_1 = \max_{1 \leq i,j \leq p} (\sum_{i=1}^p |m_{ij}|)$ is the subordinate norm to the vector norm $\|\cdot\|_1$.  

19
Proof. The proof of this lemma is given by Ciuperca and Salloum [8]. □

Lemma 2 Let the $\eta$-neighborhood of $\beta^0$, $\mathcal{V}_\eta(\beta^0) = \{\beta \in \Gamma; \|\beta - \beta^0\|_2 \leq \eta\}$, with $\eta \to 0$. Then, under assumptions (A1)-(A4), for all $\beta \in \mathcal{V}_\eta(\beta^0)$, we have

$$V_{1n}(\beta) = V_{1n}^0 + o_p(\beta - \beta^0).$$

(28)

We recall that, $V_{1n}(\beta) = \frac{1}{n\theta_n} \sum_{i \in I} \hat{g}_i(\beta)$.

Proof. By the Taylor’s expansion up to order 2 of $\hat{g}_i(\beta)$ at $\beta = \beta^0$, we obtain

$$V_{1n}(\beta) = \left[ \frac{1}{n\theta_n} \sum_{i \in I} \hat{f}_i(\beta^0)\varepsilon_i - \frac{1}{n\theta_n} \sum_{i \in I} \hat{f}_i(\beta^0)\hat{f}_i^t(\beta^0) \right]$$

$$+ \frac{1}{n\theta_n} \sum_{i \in I} \hat{M}_i \varepsilon_i - \frac{1}{n\theta_n} \sum_{i \in I} \hat{f}_i(\beta^0)\hat{f}_i(\beta^{(1)}) (\beta - \beta^0)^t$$

$$- \frac{1}{n\theta_n} \sum_{i \in I} \hat{M}_i \hat{f}_i(\beta^{(1)}) (\beta - \beta^0)^t - \frac{1}{n\theta_n} \sum_{i \in I} \hat{M}_i \hat{f}_i(\beta^0) (\beta - \beta^0)^t$$

(29)

$$+ \frac{1}{n\theta_n} \sum_{i \in I} \hat{M}_2 \hat{f}_i(\beta^0) (\beta - \beta^0)^t + \frac{1}{n\theta_n} \sum_{i \in I} \hat{M}_2 \hat{M}_3 \hat{f}_i(\beta - \beta^0)(\beta - \beta^0)^t,$$

where, $\hat{M}_i$, $\hat{M}_1i$ and $\hat{M}_2i$ are a $d \times d$ squares matrices, defined by

$$\hat{M}_i = \sum_{k=1}^d \frac{\partial f_i(\beta^{(2)})}{\partial \beta_k}(\beta_k - \beta^0_k), \hat{M}_1i = \left( \frac{\partial^2 f_i(\beta^{(3)})}{\partial \beta_k \partial \beta_l} \right)_{1 \leq k, l \leq d}$$

and

$$\hat{M}_2i = \left( \frac{\partial^2 f_i(\beta^{(4)})}{\partial \beta_k \partial \beta_l} \right)_{1 \leq k, l \leq d}.$$ Here, for $1 \leq k, l \leq d$, $\beta_k$ denotes the k-th component of $\beta$ and $\beta^{(a)}_{i,kl} = \beta^0 + u^{(a)}_{i,kl}(\beta - \beta^0)$, with $u^{(a)}_{i,kl} \in [0, 1]$ and $a \in \{2, 3, 4\}$.

For the first term of the right-hand side of (29), we have

$$\frac{1}{n\theta_n} \sum_{i \in I} \hat{f}_i(\beta^0)\varepsilon_i - \frac{1}{n\theta_n} \sum_{i \in I} \hat{f}_i(\beta^0)\hat{f}_i^t(\beta^0) = V_{1n}^0,$$

(30)

By Bienaymé-Tchebychev’s inequality and assumption (A1), we obtain that for all $C_1 > 0$ and $i \in I$

$$P[|\varepsilon_i| > C_1] \leq \frac{\sigma_i^2}{C_1},$$

(31)

For $1 \leq k, l \leq d$ and for any fixed $i$ such that $i \in I$, denote by $M_{i,kl}$ the following random variable designating the term $(j,l)$ of the matrix $\hat{M}_i$ such that

$$M_{i,kl} = \sum_{k=1}^d \frac{\partial f_i(\beta^{(2)})}{\partial \beta_k}(\beta_k - \beta^0_k).$$

Using assumption (A3), we have with a probability one that $|M_{i,kl}| \leq \|\beta - \beta^0\|_2$. Applying Lemma[1][iii], for all $C_2 > 0$ and $i \in I$, we obtain that

$$\|M_i\|_1 \leq C_2 \|\beta - \beta^0\|_2.$$

(32)
Using assumptions (A3) and relations (34), (36), for the fifth term of the right-hand side of (29), we have 
\[ \frac{1}{n \theta_{nk}} \sum_{i \in I} M_i \varepsilon_i = o_P(\beta - \beta^0). \]  
(33)

By Markov’s inequality, taking also account assumption (A4), then for any fixed 
\[ i \] and for \( 1 \leq k \leq l, \) we have that 
\[ P\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| > \epsilon_1 \right) \leq E\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| \right)/\epsilon_1. \]

For \( \epsilon_1 = E\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| / \epsilon \right), \) we obtain that 
\[ P\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| > E\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| / \epsilon \right) \right) \leq \epsilon. \]  
This last relation together with Lemma 1(i) imply 
\[ P\left( \left\| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right\|_1 \right) \leq \frac{d}{\epsilon} \max\left( E\left( \left| \frac{\partial f_i(\beta_{i,k}^{(2)})}{\partial \beta_k} \right| \right) \right) \leq \epsilon. \]  
(34)

For the third term of the right-hand side of (29), by assumption (A3) and 
using relation (34), we obtain that 
\[ (n \theta_{nk})^{-1} \sum_{i \in I} f_i(\beta^0) f_i(\beta^{(1)})(\beta - \beta^0)^t \|_1 \leq (n \theta_{nk})^{-1} \sum_{i \in I} \| f_i(\beta^0) f_i(\beta^{(1)})(\beta - \beta^0)^t \|_1. \]

Using assumption (A3), we obtain that for \( 1 \leq j, k \leq d, \) for all \( \epsilon > 0 \) there exists \( \epsilon_3 > 0 \) such that 
\[ P\left( \left| \frac{\partial^2 f_i(\beta_{i,k}^{(3)})}{\partial \beta_j \partial \beta_k} \right| \geq \epsilon_3 \right) \leq \epsilon. \]  
By Lemma 1(iii), we have that for all \( \epsilon > 0, \)

\[ P\left( \left\| M_{1i} \right\|_1 \geq \epsilon_3 \right) \leq \epsilon. \]  
(36)

Using assumption (A3) and by a similar arguments as \( M_{1i}, \) we can demonstrate that, for \( 1 \leq j, k \leq d, \) for all \( \epsilon > 0 \) there exists \( \epsilon_4 > 0 \) such that

\[ P\left( \left\| M_{2i} \right\|_1 \geq \epsilon_4 \right) \leq \epsilon. \]  
(37)

For the fourth term of the right-hand side of (29), using relations (32), (34) and 
the fact that \( \| \beta - \beta^0 \|_2 \leq \eta, \) we obtain that

\[ \frac{1}{n \theta_{nk}} \sum_{i \in I} M_i f_i(\beta^{(1)})(\beta - \beta^0)^t = o_P(\beta - \beta^0). \]  
(38)

Using assumptions (A3) and relations (34), (36), for the fifth term of the right-hand side of (29), we have

\[ \frac{1}{n \theta_{nk}} \sum_{i \in I} M_{1i} f_i(\beta^0)(\beta - \beta^0)^t = o_P(\beta - \beta^0). \]  
(39)

In the same way, using assumption (A3) and relations (34), (36), (37), we obtain that

\[ \frac{1}{n \theta_{nk}} \sum_{i \in I} (\beta - \beta^0) M_{2i} f_i(\beta^0) = o_P(\beta - \beta^0), \]  
(40)
and
\[
\frac{1}{n \theta_{nk}} \sum_{i \in I} (\beta - \beta^0) M_{2i} M_{3i} (\beta - \beta^0)^t = o_g(\beta - \beta^0),
\]
(41)

Combining relations (30), (33), (35) and (38)-(41), we obtain that \( V_{1n}(\beta) = V_{1n}^0 + o_g(\beta - \beta^0) \).

Similarly of the Lemma 2, we can demonstrate easily that \( V_{2n}(\beta) = V_{2n}^0 + o_g(\beta - \beta^0) \). Consequently, we have \( M_n(\beta) = M_n(\beta^0) + o_g(\beta - \beta^0) \), with \( M_n(\beta) \) given in Remark 1. To simplify notation we write \( M_n(\beta^0) = M_n^0 \).

The next lemma gives the behavior of \( \sum_{i \in I} z_i(\beta) \) in the neighborhood of \( \beta^0 \).

**Lemma 3** Let the \( \eta \)-neighborhood of \( \beta^0 \), \( V_\eta(\beta^0) = \{ \beta \in \Gamma; \|\beta - \beta^0\|_2 \leq \eta \} \), with \( \eta \to 0 \). Then, under assumptions (A1)-(A6), for all \( \beta \in V_\eta(\beta^0) \), we have
\[
\frac{1}{n \theta_{nk}} \sum_{i \in I} z_i(\beta) = \frac{1}{n \theta_{nk}} \sum_{i \in I} z_i(\beta^0) + \frac{1}{2} (M_n^0)^t (\beta - \beta^0) + o_g(\beta - \beta^0).
\]
(42)

**Proof.** Under assumption (A5), using Lemma 2 then the Taylor’s expansion up to the order 2 of \( (n \theta_{nk})^{-1} \sum_{i \in I} z_i(\beta) \) at \( \beta = \beta^0 \) is
\[
\frac{1}{n \theta_{nk}} \sum_{i \in I} z_i(\beta) = \frac{1}{n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) \varepsilon_i + \frac{1}{2n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) (\beta - \beta^0)f_i(\beta^0) - \frac{1}{2n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) f_i(\beta^0)^t (\beta - \beta^0)f_i(\beta^0)
\]
\[
- \frac{1}{6n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) (\beta - \beta^0)^t M_{4i} (\beta - \beta^0)
\]
\[
- \frac{1}{4n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) (\beta - \beta^0) f_i(\beta^0)^t (\beta - \beta^0)
\]
\[
- \frac{1}{12n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) (\beta - \beta^0) f_i(\beta^0)^t M_{4i} (\beta - \beta^0)
\]
\[
- \frac{1}{12n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} f_i(\beta^0) (\beta - \beta^0) (\beta - \beta^0)^t M_{4i} (\beta - \beta^0)
\]
\[
+ \frac{1}{6n \theta_{nk}} \sum_{i \in I} (M_n^0)^t (V_{1n}^0)^{-1} D_i (\beta - \beta^0)^t \varepsilon_i,
\]
(43)

where \( M_{4i} = (\frac{\partial^2 f_i(\beta^0)}{\partial \beta_k \partial \beta_l})_{1 \leq k, l \leq d} \) is a \( d \times d \) matrix and \( D_i = \left( \sum_{l=1}^{d} \sum_{k=1}^{d} \frac{\partial^2 f_i(\beta^0)}{\partial \beta_k \partial \beta_l} \cdot (\beta_k - \beta^0_k)(\beta_l - \beta^0_l) \right) \) is a vector of dimension \( (d \times 1) \). For \( 1 \leq k, l \leq d, \beta_k \) denotes the \( k \)-th component of \( \beta \) and \( \beta^{(a)}_{i,k} = \beta^0 + u^{(a)}_{i,k} (\beta - \beta^0) \), with \( u^{(a)}_{i,k} \in [0, 1] \) and \( a \in \{5, 6\} \).
For the second term of the right-hand side of (43), by the law of large numbers, the term \( \frac{1}{2n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)\varepsilon_i \) converges almost surely to the expected of \( (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)\varepsilon_i \), as \( n \to \infty \). Since \( \varepsilon_i \) is independent of \( X_i \) and \( \mathbb{E}[\varepsilon_i] = 0 \), we have

\[
\frac{1}{2n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)\varepsilon_i = o_P(\beta - \beta^0). \tag{44}
\]

For the third term of the right-hand side of (43), by a simple computation, we can obtain

\[
\frac{1}{6n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)^t M_{4i}(\beta - \beta^0) = o_P(\beta - \beta^0). \tag{45}
\]

Using assumption (A3), (A4) and by an similar arguments to those for relations (34) and (36), we have for the fourth, fifth and the sixth terms of the right-hand side of (43), respectively, that

\[
\frac{1}{4n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)^t f_i(\beta^0)(\beta - \beta^0) = o_P(\beta - \beta^0), \tag{46}
\]

\[
\frac{1}{12n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)^t M_{4i}(\beta - \beta^0) = o_P(\beta - \beta^0). \tag{47}
\]

and

\[
\frac{1}{12n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} f_i(\beta^0)(\beta - \beta^0)^t M_{4i}(\beta - \beta^0) = o_P(\beta - \beta^0). \tag{48}
\]

For any fixed \( i \), such that \( 1 \leq i \leq n\theta \) and for \( 1 \leq s \leq d \), let us denote by \( D_{is} \) the following random variable designating the \( s \)-th component of the vector \( D_i \), such that

\[
D_{is} = \sum_{l=1}^{d} \sum_{k=1}^{d} \frac{\partial^3 f_i(\beta_{k}^{(0)})}{\partial \beta_{l} \partial \beta_{k} \partial \beta_{s}} (\hat{\beta}_k - \beta_0^k)(\hat{\beta}_l - \beta_0^l). \tag{49}
\]

Applying Lemma 1 and by assumption (A3), we have for all \( C_3 > 0 \)

\[
\| D \|_1 \leq C_3 \| \beta - \beta^0 \|_2^2. \tag{50}
\]

The above equation, together with (31), implies that

\[
\frac{1}{6n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} D_{i} \varepsilon_i = o_P(\beta - \beta^0). \tag{51}
\]

Finally, for the term \( (12n\theta_{nk})^{-1} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} D_{i}(\beta - \beta^0)^t M_{4i}(\beta - \beta^0) \), assumption (A3), together with the relation (49) yield

\[
\frac{1}{12n\theta_{nk}} \sum_{i \in I} (M_n^0)^{\frac{1}{2}} (V_{1n}^0)^{-1} D_{i}(\beta - \beta^0)^t M_{4i}(\beta - \beta^0) = o_P(\beta - \beta^0). \tag{52}
\]
Combining relations (44)-(48), (50) and (51), lemma yields.

**Proof of Proposition 1** By the definition of the empirical likelihood ratio, we have the constraints $\sum_{i \in I} p_i g_i(\beta) = \sum_{j \in J} q_j g_j(\beta) = 0_d$, which give $\sum_{i \in I} p_i z_i(\beta) = \sum_{j \in J} q_j z_j(\beta) = 0_d$. Using the value of $p_i$ and $q_j$ given by (49) and (53) respectively and by an elementary calculation, we obtain

$$\sum_{i \in I} p_i z_i(\beta) = \frac{1}{n\theta nk} \sum_{i \in I} z_i(\beta) - \frac{1}{n\theta nk^2} \sum_{i \in I} \frac{z_i(\beta)z_i'(\beta)}{\theta nk} z_i(\beta) = 0_d,$$

and

$$\sum_{j \in J} q_j z_j(\beta) = \frac{1}{n(1 - \theta nk)} \sum_{j \in J} z_j(\beta) + \frac{1}{n(1 - \theta nk)^2} \sum_{j \in J} \frac{z_j(\beta)z_j'(\beta)}{\theta nk} z_j(\beta) = 0_d.$$

For the term $(n\theta nk)^{-1} \sum_{i \in I} z_i(\beta)$ of (52), by the Lemma 3 have that

$$\frac{1}{n\theta nk} \sum_{i \in I} z_i(\beta) = \frac{1}{n\theta nk} \sum_{i \in I} z_i(\beta) - \frac{1}{2}(M^0_\beta)^{\frac{1}{2}}(\beta - \beta^0) + o_p(\beta - \beta^0).$$

For the term $(n\theta nk)^{-1} \sum_{i \in I} \frac{z_i(\beta)z_i'(\beta)}{\theta nk} z_i(\beta)$, using Proposition 1 of [8], we have that for all $\epsilon > 0$, there exists $N_1, N_2 > 0$, such that

$$\mathbb{P} \left[ \frac{1}{N_1} \sum_{i \in I} z_i(\beta)z_i'(\beta) \leq \sum_{i \in I} \frac{z_i(\beta)z_i'(\beta)}{1 + \frac{\lambda(\beta)}{\theta nk}} z_i(\beta) \leq \frac{1}{N_2} \sum_{i \in I} z_i(\beta)z_i'(\beta) \right] < \epsilon.$$

This implies that, in order to study the second term of the right-hand side of (52), we must study only $(n\theta nk)^{-1} \sum_{i \in I} z_i(\beta)z_i'(\beta)$.

By Lemma 3 we have that $V_{1n}(\beta) = V_{1n}^0 + o_p(\beta - \beta^0)$. Under assumption (A5), the Taylor’s expansion up the order 2 of $z_i(\beta)$ at $\beta = \beta^0$ is

$$z_i(\beta) = (M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}f_i(\beta^0)\varepsilon_i + \frac{1}{2}(M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}M_{6i}(\beta - \beta^0)\varepsilon_i,$$

$$-\frac{1}{2}(M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}f_i(\beta^0)f_i'(\beta^0)(\beta - \beta^0),$$

$$-\frac{1}{6}(M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}f_i(\beta^0)(\beta - \beta^0)^4M_{7i}(\beta - \beta^0),$$

$$-\frac{1}{4}(M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}M_{6i}(\beta - \beta^0)f_i'(\beta^0)(\beta - \beta^0),$$

$$-\frac{1}{12}(M^0_\beta)^{\frac{1}{2}}(V^0_{1n})^{-1}M_{6i}(\beta - \beta^0)(\beta - \beta^0)^4M_{7i}(\beta - \beta^0).$$

where, $M_{6i}$ and $M_{7i}$ are a $d \times d$ squares matrices, defined by $M_{6i} = \left(\frac{\partial^2 f_i(\beta^0)}{\partial^2 \beta_i \partial^2 \beta_j}\right)_{1 \leq k, l \leq d}$

and $M_{7i} = \left(\frac{\partial^2 f_i(\beta^0)}{\partial^2 \beta_i \partial^2 \beta_j} \right)_{1 \leq k, l \leq d}$ and $\beta^{(a)} = \beta^0 + u_{i, kl}(\beta - \beta^0)$, with $u_{i, kl} \in [0, 1]$ and $a \in \{7, 8\}$.  

24
For the second term of (52), using the Taylor’s expansion of $z_i(\beta)$ in a neighborhood of $\beta^0$ given by the relation (55) and with a similar argument to the one used in Lemma 3 for the first term of (52), together with the assumptions (A3), (A4), we obtain

$$\frac{1}{n\theta_{nk}} \sum_{i \in I} z_i(\beta)z_i'(\beta) = \frac{1}{n\theta_{nk}} \sum_{i \in I} z_i(\beta^0)z_i'(\beta^0) + \Delta_n^0 + o_{\theta}(\beta - \beta^0).$$

(56)

where $\Delta_n^0 = (n\theta_{nk})^{-1} \sum_{i \in I}(M_i^0)\hat{\gamma} V_i(\theta_{nk})^{-1} f_i(\beta^0) f_i^t(\beta - \beta^0)[(M_i^0)\hat{\gamma} V_i^0(\theta_{nk})^{-1} f_i(\beta - \beta^0)] f_i^t(\beta^0)$.

In the same way, for the observations $j \in J$, we obtain

$$\frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_j(\beta) = \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_j(\beta^0) - \frac{1}{2}(M_n^0)\hat{\gamma} (\beta - \beta^0) + o_{\theta}(\beta - \beta^0),$$

(57)

and

$$\frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_i(\beta)z_j'(\beta) = \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_j(\beta^0)z_j'(\beta^0) + \Delta_n^2 + o_{\theta}(\beta - \beta^0),$$

(58)

where $\Delta_n^2 = (n(1 - \theta_{nk}))^{-1} \sum_{j \in J}(M_j^0)\hat{\gamma} V_j(\theta_{nk})^{-1} f_j(\beta^0) f_j^t(\beta - \beta^0)[(M_j^0)\hat{\gamma} V_j^0(\theta_{nk})^{-1} f_j(\beta - \beta^0)] f_j^t(\beta^0)$.

To facilitate writing, we consider the $d \times d$ squares matrices, defined by

$$S_n^0 = \frac{1}{n\theta_{nk}} \sum_{i \in I} z_i(\beta^0)z_i'(\beta^0) + \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_j(\beta^0)z_j'(\beta^0),$$

$$\Delta_n = \Delta_n^1 + \Delta_n^2,$$

and we define the vector

$$\psi_n^0 = \frac{1}{n\theta_{nk}} \sum_{i \in I} z_i(\beta^0) - \frac{1}{n(1 - \theta_{nk})} \sum_{j \in J} z_j(\beta^0).$$

On the other hand, we have $\Phi_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) = 0_d$. Using relations (42) and (56)-(58), we obtain $[(n\theta_{nk})^{-1} \sum_{i \in I} z_i(\beta^0) - \frac{1}{2}(M_i^0)\hat{\gamma} (\beta(\theta_{nk}) - \beta^0) - \Delta_n^1 \lambda(\theta_{nk}) - (n\theta_{nk})^{-1} \sum_{i \in I} z_i(\beta^0)z_i(\beta^0)\hat{\lambda}(\theta_{nk}) - (n\theta_{nk})^{-1} \sum_{j \in J} z_j(\beta^0) - \frac{1}{2}(M_j^0)\hat{\gamma} (\beta(\theta_{nk}) - \beta^0) + \Delta_n^2 \lambda(\theta_{nk}) + (n(1 - \theta_{nk}))^{-1} \sum_{j \in J} z_j(\beta^0)z_j(\beta^0)\hat{\lambda}(\theta_{nk})] = 0_d$. Using the notations given above, then we obtain

$$\lambda(\theta_{nk}) = (S_n^0 + \Delta_n^0)^{-1}\psi_n^0 + o_{\theta}(\beta(\theta_{nk}) - \beta^0).$$

(59)

The limited development of the statistic $Z_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk}))$ specified by the relation (12) in the neighborhood of $(\lambda, \beta) = (0_d, \beta^0)$ up to order 2, can
be written

\[
Z_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) = \left[2\lambda^t(\theta_{nk}) \left( \frac{1}{\theta_{nk}} \sum_{i \in I} z_i(\beta^0) - \frac{1}{1 - \theta_{nk}} \sum_{j \in J} z_j(\beta^0) \right) \right] \\
- \left[ \hat{\lambda}^t(\theta_{nk}) \left( \frac{1}{1 - \theta_{nk}} \sum_{i \in I} g_i(\beta^0) g_i^t(\beta^0) + \frac{1}{(1 - \theta_{nk})^2} \sum_{j \in J} g_j(\beta^0) g_j^t(\beta^0) \right) \hat{\lambda}(\theta_{nk}) \right] \\
+ \left[ 2\lambda^t(\theta_{nk}) \left( (M^0_n)_{1}^t (V^0_{1n})^{-1} \frac{1}{\theta_{nk}} \sum_{i \in I} g_i(\beta^0) - (M^0_n)_{1}^t (V^0_{2n})^{-1} \frac{1}{1 - \theta_{nk}} \sum_{j \in J} g_j(\beta^0) \right) \hat{\beta}(\theta_{nk}) - \beta^0 \right] \\
- \frac{2\lambda^t(\theta_{nk})}{1 - \theta_{nk}} \sum_{j \in J} g_j(\beta^0) \frac{\partial (M^1_n(\beta)(V_{1n}(\beta))^{-1})}{\partial \beta} \\
+ \frac{1}{31} \left[ T_1 + 3T_2 + 3T_3 + T_4 \right],
\]

(60)

where,

\[
T_1 = \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{i \in I} \frac{\partial^2 Z_{nk}(\gamma_0, \lambda_{jkl}^{(i)}, \lambda_{jkl}^{(i)})}{\partial \lambda_{jkl} \partial \lambda_{jkl} \partial \beta} (\hat{\beta}_j - \beta_j^0)(\hat{\beta}_k - \beta_k^0)(\hat{\beta}_l - \beta_l^0),
T_2 = \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{i \in I} \frac{\partial^2 Z_{nk}(\gamma_0, \lambda_{jkl}^{(i)}, \lambda_{jkl}^{(i)})}{\partial \lambda_{jkl} \partial \lambda_{jkl} \partial \beta} (\hat{\lambda}_j)(\hat{\lambda}_k)(\hat{\lambda}_l - \beta_l^0),
T_3 = \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{i \in I} \frac{\partial^2 Z_{nk}(\gamma_0, \lambda_{jkl}^{(i)}, \lambda_{jkl}^{(i)})}{\partial \lambda_{jkl} \partial \lambda_{jkl} \partial \lambda_{jkl}} (\hat{\lambda}_j)(\hat{\lambda}_k - \beta_k^0)(\hat{\lambda}_l - \beta_l^0),
T_4 = \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{i \in I} \frac{\partial^2 Z_{nk}(\gamma_0, \lambda_{jkl}^{(i)}, \lambda_{jkl}^{(i)})}{\partial \lambda_{jkl} \partial \lambda_{jkl} \partial \lambda_{jkl}} (\hat{\lambda}_j)(\hat{\lambda}_k)(\hat{\lambda}_l),
\]

and for \( 1 \leq j \leq d \), \( \hat{\beta}_j \) is the j-th component of \( \hat{\beta}(\theta_{nk}) \), \( \hat{\lambda}_j \) is the j-th component of \( \hat{\lambda}(\theta_{nk}) \). For all \( 1 \leq j, k, l \leq d \), \( \lambda_{jkl}^{(i)} = u_{jkl}^{(i)}(\beta - \beta^0) \) and \( \beta_{jkl}^{(i)} = \beta^0 + v_{jkl}^{(i)}(\beta - \beta^0) \), with \( u_{jkl}^{(i)}, v_{jkl}^{(i)} \in [0, 1] \) and \( b \in \{1, 2, 3, 4\} \).

We note that, the derivatives \( \partial (M^1_n(\beta)(V_{1n}(\beta))^{-1})/\partial \beta \) and \( \partial (M^1_n(\beta)(V_{2n}(\beta))^{-1})/\partial \beta \) are considered term by term.

Now, we replace \( \hat{\lambda}(\theta_{nk}) \) in the relation (60) by their value obtained in (59). For the first term of (60), using notations given above, we find that this term is equal to \( 2(n^0_n)^t(S_n^0 + \Delta_n^0)^{-1}\psi_n^0 \).

Similarly, the second term of (60) is \( n(\psi_n^0)^t(S_n^0 + \Delta_n^0)^{-1}\psi_n^0 \). We know that \( V_{1n}^0 = (n\theta_{nk})^{-1} \sum_{i \in I} g_i(\beta^0) \) and \( V_{2n}^0 = (n(1 - \theta_{nk}))^{-1} \sum_{j \in J} g_j(\beta^0) \). Then the third term of (60) converge almost surely to zero, as \( n \rightarrow \infty \).

By the central limit theorem, we have that \( (n\theta_{nk})^{-1} \sum_{i \in I} g_i(\beta^0) = o_p(n(\theta_{nk}))^{-1/2} \) and \( (n(1 - \theta_{nk}))^{-1} \sum_{j \in J} g_j(\beta^0) = o_p((n(1 - \theta_{nk}))^{-1/2}) \). Then, the fourth term of (60) is \( o_p(n^0_n)^t(S_n^0 + \Delta_n^0)^{-1}\psi_n^0 \).

For the last term of (60), using assumptions (A2)-(A4) and by elementary cal-
culations, we prove that this term is $\xi(\theta_{nk})$, where

$$
\xi(\theta_{nk}) = o_{\mathcal{P}}(\|\hat{\beta}(\theta_{nk}) - \beta^0\|_2) + o_{\mathcal{P}}(\|\hat{\lambda}(\theta_{nk})\|_2) + o_{\mathcal{P}}(\|\hat{\lambda}(\theta_{nk})\|_2\|\hat{\beta}(\theta_{nk}) - \beta^0\|_2).
$$

Combining the obtained results, we obtain

$$
Z_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) = n(\psi_n^0)^t(S_n^0 + \Delta_n^0)^{-1}\psi_n(1 + o_{\mathcal{P}}(1)) + \xi(\theta_{nk}).
$$

We can see that, when $\beta - \beta^0 = O_d$, i.e., $\beta = \beta^0$, $n(\psi_n^0)^t(S_n^0 + \Delta_n^0)^{-1}\psi_n$ achieves its maximum in the neighborhood of $\beta^0$. This means when $n \to \infty$, $Z_{nk}(\lambda, \gamma_0, \beta)$ has one local maximum in any $\eta$-neighborhood of $\beta^0$ and then $\hat{\beta}(\theta_{nk}) - \beta_0 = O_{\mathcal{P}}(1)$.

For $S_n^0$, by the law of large numbers, the terms $(n\theta_{nk}^2)^{-1}\sum_{i \in I} z_i(\beta^0)z_i^t(\beta^0)$ and $(n(1 - \theta_{nk})^2)^{-1}\sum_{i \in I} z_i(\beta^0)z_i^t(\beta^0)$ converge almost surely to the expected of $z_i(\beta^0)z_i^t(\beta^0)$ and $z_i(\beta^0)z_i^t(\beta^0)$, respectively as $n \to \infty$. Using this fact and the fact that $-V_{n^0} \overset{a.s.}{\to} V$ and $-V_{2n} \overset{a.s.}{\to} V$, by the law of large numbers, we can demonstrate easily that $S_n^0 = \mathbf{I}_d + O_{\mathcal{P}}(n^{-1/2})$. Using this fact and relation \[59\], we obtain that $\hat{\lambda}(\theta_{nk}) = \psi_n^0 + O_{\mathcal{P}}(n^{-1/2})$. By the central limit theorem and the fact that $E[\hat{g}_i(\beta^0)] = 0$ for $i = 1, \ldots, n$, each term of $\psi_n^0$ is $O_{\mathcal{P}}(n^{-1/2})$, which implies $\hat{\lambda}(\theta_{nk}) = O_{\mathcal{P}}(n^{-1/2})$. The lemma is completely proved.

**Proof of Theorem** \ref{thm1} Using the proof of Proposition \ref{prop1} we have $\hat{\lambda}(\theta_{nk}) = \psi_n^0 + O_{\mathcal{P}}(n^{-1/2})$ and $S_n^0 = \mathbf{I}_d + O_{\mathcal{P}}(n^{-1/2})$. By the Linderberg-Feller theorem, we have that $\sqrt{n}\hat{\lambda}(\theta_{nk}) \sim \mathcal{N}(0, \mathbf{I}_d)$. Then, using also relation \[59\], we obtain

$$
Z_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) = n(\psi_n^0)^t(S_n^0)^{-1}\psi_n^0 + O_{\mathcal{P}}(n^{-1/2})
$$

where $\xi(\theta_{nk})$ is given by the relation \[61\]. Then, since $\sqrt{n}\hat{\lambda}(\theta_{nk}) \sim \mathcal{N}(0, \mathbf{I}_d)$, we deduce

$$
Z_{nk}(\gamma_0, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})) \overset{\mathcal{L}}{\to} \chi^2(d).
$$

The theorem is proved.

**Proof of Proposition** \ref{prop2} For each method, the proof is similar to the proof of Proposition \ref{prop1}.

**Proof of Theorem** \ref{thm2} Using Proposition \ref{prop2} the proof of this theorem is similar to the proof of Theorem \ref{thm1} for the three proposed methods.
References

[1] Bai, J. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics, 91*, 299-323, 1999.

[2] Boldea, O., Hall, A.R. Estimation and inference in unstable nonlinear least squares models. *Journal of Econometrics, 172*(1), 158-167, 2013.

[3] Chen, S. On the accuracy of empirical likelihood confidence region for linear regression model, *Ann. Inst. Statistic. Math.*, 45, 621-637, 1993.

[4] Chen, S. Empirical likelihood confidence interval for linear regression coefficients, *Ann. Inst. Statistic. Math.*, 45, 621-637, 1994.

[5] Chen, S., Cui, H. An extended empirical likelihood for generalized linear models, *Statistica Sinicia*, 13, 69-81, 2003.

[6] Ciuperca, G. A general criterion to determine the number of change-points. *Statistics and Probability Letters, 81*(8), 1267-1275, 2011.

[7] Ciuperca G. Empirical likelihood for nonlinear model with missing responses, *Journal of Statistical Computation and Simulation, 83*(4), 737-756, 2013.

[8] Ciuperca, G., Salloum, Z. Empirical likelihood test in a posteriori change-point nonlinear model. *arXiv:1312.3757*, 2013.

[9] Horvitz, DG., Thompson, DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat assoc, 47*, 663-685, 1952.

[10] Jing, B. Two-sample empirical likelihood method. *Statistics and Probability Letters, 24*, 315-319, 1995.

[11] Kim, H.j., Siegmund, D. The likelihood ratio test for a change-point in simple linear regression model. *Biometrika, 76*, 409-423, 1989.

[12] Kolaczyk, E. D. Empirical likelihood for generalized linear models. *Statistica Sinicia, 4*, 199-218, 1994.

[13] Little, R.J.A, Rubin, D.B. Statistical analysis with missing data. *John Wiley and sons, Inc., New York*, 1987.

[14] Liu, Y., Zou, C., Zhang, R. Empirical likelihood for the two-sample mean problem. *Statistics and Probability Letters, 78*, 548-556, 2008.

[15] Liu, Y., Zou, C., Zhang, R. Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics-Theory and Methods, 37*, 2551-2563, 2008.

[16] Muller, U.U. Estimating linear functionals in nonlinear regression with responses missing at random. *Annal statistic, 37*(5A), 2245-2277, 2009.
[17] Ning, W., Pailden, J., Gupta, A. Empirical likelihood ratio test for the epidemic change model. *Journal of Data Science*, **10**, 107-127, 2012.

[18] Owen, A. B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249, 1988.

[19] Owen, A. B. Empirical likelihood ratio confidence regions. *Annal statistic*, **18**, 90-120, 1990.

[20] Owen, A. B. Empirical likelihood for linear model. *Annal statistic*, **19**, 1725-1747, 1991.

[21] Qin, Y., Li, L., Lei, Q. Empirical likelihood for linear regression model with missing responses. *Statistics and Probability Letters*, **79**, 1391-1396, 2009.

[22] Seber G., Wild C. Testing for structural change in regression quantiles, Wiley series in probability and mathematical Statistics, Wiley, Hoboken, NJ, 2003.

[23] Sun, Z., Wang, Q., Dai, P. model checking for partially linear models with missing responses at random. *Journal of Multivariate Analysis*, **100**, 636-651, 2009.

[24] Wei, Yu., Cuizhen, Ni., Wangli, Xu. An empirical likelihood inference for the coefficient difference of a two-sample linear model with missing response data. *Metrika*, **81**(8), 1267-1275, 2013.

[25] Xue, L. Empirical likelihood for linear models with missing responses. *Journal of Multivariate Analysis*, **100**, 1353-1366, 2009.

[26] Zi, X., Zou, C., Liu, Y. Two-sample empirical likelihood method for difference between coefficients in linear regression model. *Stat Papers*, 2010.

[27] Zou C., Liu Y., Qin P., Wang Z. Empirical likelihood ratio test for a change point. *Statistics and Probability Letters*, **77**, 374-382, 2007.