

Research Article

The Empirical Cressie-Read Test Statistics for Longitudinal Generalized Linear Models

Junhua Zhang,¹ Ruiqin Tian,² Suigen Yang,^{2,3} and Sanying Feng²

¹ College of Mechanical Engineering, Beijing Information Science and Technology University, Beijing 100192, China

² College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

³ College of Sciences, Tianjin University of Commerce, Tianjin 300134, China

Correspondence should be addressed to Ruiqin Tian; tianruiqin@emails.bjut.edu.cn

Received 10 July 2013; Revised 19 December 2013; Accepted 24 January 2014; Published 18 March 2014

Academic Editor: Jen-Tzung Chien

Copyright © 2014 Junhua Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For the marginal longitudinal generalized linear models (GLMs), we develop the empirical Cressie-Read (ECR) test statistic approach which has been proposed for the independent identically distributed (i.i.d.) case. The ECR test statistic includes empirical likelihood as a special case. By adopting this ECR test statistic approach and taking into account the within-subject correlation, the efficiency theory results of estimation and testing based on ECR are established under some regularity conditions. Although a working correlation matrix is assumed, there is no need to estimate the nuisance parameters in the working correlation matrix based on the quadratic inference function (QIF). Therefore, the proposed ECR test statistic is asymptotically a standard χ^2 limit under the null hypothesis. It is shown that the proposed method is more efficient even when the working correlation matrix is misspecified. We also evaluate the finite sample performance of the proposed methods via simulation studies and a real data analysis.

1. Introduction

Longitudinal studies are increasingly common in many scientific research fields, including epidemiology, econometrics, medicine, and life and social sciences. For example, longitudinal studies are often used in psychology to study developmental trends across the life span, in sociology to study life events throughout lifetimes or generations, and in medical area to take different treatments at the start of the study and to see what kind of effects the assigned patients have with each treatment by week or by year.

Longitudinal data modeling is a statistical method often used in the experiments that are designed such that responses on the same experimental units are observed at each repetition. However, generalized linear models (GLMs [1]) have become a favored tool for the modelling of clustered and longitudinal data by allowing for the non-Gaussian data and nonlinear link functions, such as binomial or Poisson's type responses. The intrinsic complexity of longitudinal GLMs makes them challenging. First, one characteristic of longitudinal data is the within-subject correlation among the repeated measurements. Ignoring this within-subject correlation causes a loss of efficiency in general problems, and

the presence of correlation makes it hard to establish the underlying asymptotic theory. Second, the full likelihood for longitudinal data is often difficult to specify, particularly for the correlated non-Gaussian data. Researchers had developed appropriate methods and criteria for longitudinal GLMs, where generalized estimating equations (GEE [2, 3]) or a recently proposed related method based on quadratic inference functions (QIF; see [4]) is an attractive option for longitudinal regression, particularly when marginal relationships rather than within-subject correlation are of primary interest.

In the present paper, we consider the popular marginal longitudinal GLMs. Suppose that $\mathbf{Y}_i = (y_{i1}, \dots, y_{im_i})^T$ is the multivariate response for the i subject and $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})^T$ is the $m_i \times p$ matrix of the covariates for the i th subject ($i = 1, \dots, n$). Observations from different subjects are independent, but those from the same subjects are correlated. Assume that the marginal mean of y_{ij} is

$$E(y_{ij} | \mathbf{x}_{ij}) = h(\boldsymbol{\beta}^T \mathbf{x}_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (1)$$

where $h(\cdot)$ is a known link function, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the unknown parameter vector of interest, and $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ is a $p \times 1$ vector for $j = 1, \dots, m_i$. In order to simplify notation, but without loss of generality, we assume $m_i \equiv m$. Recently, there has been considerable interest of investigating the GLMs, such as Liang and Zeger [2], Zeger and Liang [3], Fitzmaurice [5], Qu et al. [4], Pan [6], Balan and Schiopu-Kratina [7], Qu and Li [8], Wang and Qu [9], Wang [10], and Li et al. [11].

The empirical likelihood method was introduced by Owen [12–14] as a nonparametric method of inference based on a data-driven likelihood ratio function in the i.i.d. case and had been applied to various statistical models. For example, Kolaczyk [15] and Chen and Cui [16] made an extension to the generalized linear models. As for other studies, see Baggerly [17], DiCiccio et al. [18], Chen and Qin [19], Qin and Lawless [20], Zhu and Xue [21], and Li et al. [22] among others. Owen [23] is one of the best references for this field. Although the empirical likelihood method had been investigated by many authors, those researches had mostly focused on independent observations. For the analysis of longitudinal data, Xue and Zhu [24] applied the empirical likelihood approach to the varying coefficient models with longitudinal data, but they did not consider the within-subject correlation structure of the longitudinal data. Li et al. [25] and Li et al. [26] proposed the generalized empirical likelihood method for longitudinal data by introducing the known within-subject correlation structure. Wang et al. [27] proposed two generalized empirical likelihood methods for the longitudinal linear model by taking into consideration within-subject correlations. They showed that, by plugging the estimator of the working correlation matrix into the empirical log-likelihood ratio (ELR), the resulting ELR is asymptotically a weighted sum of independent χ_1^2 -variables with unknown weights. Thus, the empirical likelihood method needs to be adjusted so that it can be efficiently used to accommodate the correlation inherent in longitudinal data. However, in many applications, how to estimate the working correlation matrix effectively is a challenging problem.

Instead of the empirical likelihood ratio statistic, Baggerly [17] used the empirical Cressie-Read (ECR) test statistic and showed that it is also asymptotically a Chi-squared distribution in the i.i.d. case. Bravo [28] extended the ECR approach to inference for α -mixing processes. The empirical Cressie-Read test statistic has user-specified parameter $\lambda \in (-\infty, \infty)$ and encompasses several commonly used testing statistics as special cases, such as the empirical likelihood statistic ($\lambda = 0$), the maximum entropy, minimum information or the Kullback-Leibler statistic ($\lambda = -1$), the Neyman-modified χ^2 statistic or the Euclidean likelihood statistic ($\lambda = -2$), the Freeman-Tukey statistic ($\lambda = -1/2$), and Pearson's χ^2 statistic ($\lambda = 1$), the first two being defined in a limiting sense. Thus, we can use these results to construct confidence regions of the parameter $\boldsymbol{\beta}$ of interest.

In this paper we develop Baggerly's results on the ECR test statistic for the longitudinal data by combining the idea of QIF in Qu et al. [4] and apply the proposed method to the marginal longitudinal GLMs (1). Specifically, we derive the asymptotic distributions of the maximum empirical

Cressie-Read likelihood estimator (MECRLE) and the ECR test statistic. Although the working correlation matrix is assumed, the proposed method needs not to estimate the nuisance parameters associated with correlations. This is because the inverse of the commonly used working correlation matrix can be exactly represented or approximated by a linear combination of basis matrices such that the dimension of the unbiased estimating functions is greater than the number of unknown parameters. Thus, the efficiency results can be obtained by combining the ECR method with the generalized method of moments (GMM) proposed by Hansen [29]. Therefore, the proposed method inherits the advantages of the empirical likelihood, QIF, and GMM methods, and the proposed method represents a good alternative in this area as well.

The paper is organized as follows. In Section 2, we propose the more generalized method of the ECR test statistic for marginal longitudinal GLMs by combining with the QIF approach. The technical conditions and some asymptotic properties are given in Section 3. In Section 4, we present the results for simulation studies and a real dataset to illustrate the proposed methods. The technical proofs of the main results are presented in Section 5.

2. Model and ECR Test Statistic

Suppose that the population (\mathbf{X}, \mathbf{Y}) comes from the marginal longitudinal generalized linear model (1). Then the marginal mean of y_{ij} is

$$\mu_{ij} = E(y_{ij} | \mathbf{x}_{ij}) = h(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \quad (2)$$

and the marginal variance of y_{ij} is

$$\text{Var}(y_{ij} | \mathbf{x}_{ij}) = \phi \nu(\mu_{ij}), \quad (3)$$

where $\nu(\cdot)$ is a variance function and ϕ is a scale parameter.

With assumptions on the first marginal moments, the GEE [2] estimator of $\boldsymbol{\beta}$ is defined as the solution of

$$\sum_{i=1}^n \dot{\boldsymbol{\mu}}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \quad (4)$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{im})^T$, $\dot{\boldsymbol{\mu}}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$, is an $m \times p$ matrix and $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ is a working covariance matrix with \mathbf{A}_i being the $m \times m$ diagonal matrix of marginal variances $\text{Var}(\mathbf{Y}_i)$ and $\mathbf{R}(\boldsymbol{\alpha})$ as the working correlation matrix, where $\boldsymbol{\alpha}$ is a vector which fully characterizes $\mathbf{R}(\boldsymbol{\alpha})$. The main advantage of the GEE method is that it yields a consistent estimator even if the working correlation matrix is misspecified.

Qu et al. [4] introduced the quadratic inference function (QIF) by assuming that the inverse of the working correlation can be approximated by a linear combination of several basis matrices; that is,

$$\mathbf{R}^{-1} \approx a_1 T_1 + a_2 T_2 + \dots + a_k T_k, \quad (5)$$

where T_1 is the identity matrix, T_2, \dots, T_k are symmetric basis matrices which are determined by the structure of $\mathbf{R}(\boldsymbol{\alpha})$,

and a_1, \dots, a_k are constant coefficients. The advantage of this approach is that it does not require estimation of linear coefficients a_i 's which can be viewed as nuisance parameters.

To implement the QIF approach, we need to choose the basis for the inverse of the correlation matrix $\mathbf{R}(\alpha)$. Qu et al. [4] and Qu and Song [30] found that, if $\mathbf{R}(\alpha)$ is the exchangeable working correlation matrix, $\mathbf{R}^{-1} = a_1 T_1 + a_2 T_2$, where T_1 is the identity matrix and T_2 is a matrix with 0 on the diagonal and 1 off diagonal. If the working correlation matrix is AR(1) with $R_{ij} = \alpha^{|i-j|}$, then \mathbf{R}^{-1} can be written as a linear combination of three basis matrices, that is, $a_1 T_1^* + a_2 T_2^* + a_3 T_3^*$, where T_1^* is the identity matrix, T_2^* can be 1 on the subdiagonal and 0 elsewhere, and T_3^* can be 1 on the corners (1, 1) and (m, m) and 0 elsewhere. However, T_3^* can often be dropped out of the model, as removing T_3^* does not affect the efficiency of the estimator too much, and this can simplify the estimation procedure. More details can be found in Qu et al. [4], Qu and Song [30], and Qu and Li [8] among others. In addition, Qu and Lindsay [31] developed an adaptive estimating equation approach to find a reliable approximation to the inverse of the variance matrix.

Let

$$g_i(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{\mu}_i^T \mathbf{A}_i^{-1/2} T_1 \mathbf{A}_i^{-1/2} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \\ \boldsymbol{\mu}_i^T \mathbf{A}_i^{-1/2} T_2 \mathbf{A}_i^{-1/2} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \boldsymbol{\mu}_i^T \mathbf{A}_i^{-1/2} T_k \mathbf{A}_i^{-1/2} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \end{pmatrix}. \quad (6)$$

It is easy to check that the GEE defined in (4) becomes a linear combination of the above extended score vector of $\sum_{i=1}^n g_i(\boldsymbol{\beta})$. Note that the dimension of $g_i(\boldsymbol{\beta})$ is $r = kp$, and it is greater than the number of unknown parameters; thus the GEE method is unavailable. Therefore, Qu et al. [4] proposed the quadratic inference function by extending the method of GMM proposed by Hansen [29] to obtain the estimator of $\boldsymbol{\beta}$.

In this paper, we consider an alternate method using the following empirical Cressie-Read test statistics [17, 32]:

$$CR(\lambda) = \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left\{ (np_i)^{-\lambda} - 1 \right\}, \quad -\infty < \lambda < \infty, \quad (7)$$

where λ is a user-specified parameter. ECR (7) is a generalization of the empirical log-likelihood ratio statistic and can also be defined as

$$CR(\lambda) = \begin{cases} \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left\{ (np_i)^{-\lambda} - 1 \right\}, & \text{for } \lambda \neq 0, -1, \\ -2 \sum_{i=1}^n \log(np_i), & \text{for } \lambda = 0, \\ 2n \sum_{i=1}^n p_i \log(np_i), & \text{for } \lambda = -1. \end{cases} \quad (8)$$

Let p_i be a multinomial likelihood supported on the observed data, and suppose that we are interested in testing the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is the true parameter vector. The empirical Cressie-Read family of test statistics for

H_0 can be based on the following constrained minimization problem:

$$CR_\lambda(\boldsymbol{\beta}) = \min \left\{ CR(\lambda) \mid \sum_{i=1}^n p_i = 1, p_i \geq 0, \sum_{i=1}^n p_i g_i(\boldsymbol{\beta}) = 0 \right\}. \quad (9)$$

As noted by Baggerly [17], a unique value for the right-hand side of (9) exists, provided that zero is inside the convex hull of the $g_1(\boldsymbol{\beta}), \dots, g_n(\boldsymbol{\beta})$ for given $\boldsymbol{\beta}$. Applying the Lagrange multiplier method, we can obtain the solution of $\boldsymbol{\beta}$ by minimizing (9). Let

$$G = \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left\{ (np_i)^{-\lambda} - 1 \right\} + \kappa_1 \left(\sum_{i=1}^n p_i - 1 \right) + n\kappa_2^T \sum_{i=1}^n p_i g_i(\boldsymbol{\beta}), \quad (10)$$

where κ_1 and $\kappa_2 = (\kappa_1, \dots, \kappa_{kp})^T$ are the Lagrange multipliers. Let $\partial G / \partial p_i = 0$; we have

$$p_i = \begin{cases} \frac{1}{n} \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)}, & (\lambda \neq -1), \\ s \exp \left\{ t^T g_i(\boldsymbol{\beta}) \right\}, & (\lambda = -1), \end{cases} \quad (11)$$

where s and the vector $t = (t_1, \dots, t_{kp})^T$ are determined by

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)} &= 1, \\ \frac{1}{n} \sum_{i=1}^n \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)} g_i(\boldsymbol{\beta}) &= 0. \end{aligned} \quad (12)$$

In the present paper, we only are interested in the case of $\lambda \neq -1$. When $\lambda \rightarrow -1$, the estimator of $\boldsymbol{\beta}$ is viewed as the empirical exponential tilting estimator (see [33, 34]). When $\lambda \neq -1$ and substituting (11) into (7), we obtain the following empirical Cressie-Read test statistic:

$$CR_\lambda(\boldsymbol{\beta}) = \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left[\left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{\lambda/(1+\lambda)} - 1 \right]. \quad (13)$$

Baggerly [17] had applied the empirical Cressie-Read test statistic to the general parameter case and shown that all members of the empirical Cressie-Read family have a Chi-squared calibration and enjoy the advantages of the empirical likelihood method. From the proof of Theorem 1 in this paper, we can obtain that

$$\begin{aligned} CR_\lambda(\boldsymbol{\beta}) &= n \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) + o_P(1), \\ s &= \frac{1}{2} \lambda (1+\lambda) \bar{g}_n(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) + O_P(n^{-1/2} \zeta_n), \\ t(\boldsymbol{\beta}) &= (1+\lambda) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) + \zeta_n, \quad \zeta_n = o_P(n^{-1/2}), \end{aligned} \quad (14)$$

where $\bar{g}_n(\boldsymbol{\beta}) = (1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta})$ and $\widehat{\Sigma} = (1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta})$. If $\hat{\boldsymbol{\beta}}$ satisfies $CR_\lambda(\hat{\boldsymbol{\beta}}) = \inf_{\boldsymbol{\beta} \in \mathcal{B}} CR_\lambda(\boldsymbol{\beta})$, the estimator $\hat{\boldsymbol{\beta}}$ is called the maximum empirical Cressie-Read likelihood estimator (MECRLE) of the parameter $\boldsymbol{\beta}$.

3. Asymptotic Properties

To establish the asymptotic properties, we need the following regularity conditions.

- (C1) The matrix $\widehat{\Sigma} = (1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta})$ converges a.s. to an invertible positive definite matrix $\Sigma(\boldsymbol{\beta})$. This condition holds based on the weak law of large number when n tends to infinity and the cluster size is fixed.
- (C2) The domain \mathcal{B} is a compact subset of \mathbb{R}^p and the true parameter value $\boldsymbol{\beta}_0$ is in its interior. There exists a unique $\boldsymbol{\beta}_0 \in \mathcal{B}$ satisfying mean zero model assumption $E\{g_1(\boldsymbol{\beta}_0)\} = 0$.
- (C3) Let $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{im})^T$ where $\mu_{ij} = E(y_{ij} | \mathbf{x}_{ij})$; then $\boldsymbol{\mu}_i$ is a.s. continuously differentiable in $\boldsymbol{\beta}$. Denote this $m \times p$ derivative matrix by $\dot{\boldsymbol{\mu}}_i$; then $\dot{\boldsymbol{\mu}}_i$ has full column rank p a.s.
- (C4) Assume that $\partial g_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ is continuous in a neighborhood of the true value $\boldsymbol{\beta}_0$ and the rank of $E(\partial g_1(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta})$ is p . In addition, $(\partial \widehat{g}_n/\partial \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}})$ converges in probability to $E[(\partial g_1/\partial \boldsymbol{\beta})(\boldsymbol{\beta}_0)]$ when $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$.

Conditions (C1)–(C4) are actually quite mild and can be easily satisfied, and these conditions are also found in Qin and Lawless [20] and Qu et al. [4]. Conditions (C1) and (C4) ensure that the asymptotic variance exists for the proposed estimator. Condition (C2) ensures that there exists a \sqrt{n} -consistent solution in the compact subset \mathcal{B} . (C3) is a common condition used in GLMs with longitudinal data.

Theorem 1. Assume that conditions (C1–C4) hold. Then, as $n \rightarrow \infty$, with probability tending to 1, $CR_\lambda(\boldsymbol{\beta})$ attains its minimum value at some point $\widehat{\boldsymbol{\beta}}$ in the interior of the ball $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq n^{-1/3}$ and $\widehat{\boldsymbol{\beta}}$, \widehat{s} , and $\widehat{t} = t(\widehat{\boldsymbol{\beta}})$ satisfy

$$Q_{1n}(\widehat{\boldsymbol{\beta}}, \widehat{s}, \widehat{t}) = 0, \quad Q_{2n}(\widehat{\boldsymbol{\beta}}, \widehat{s}, \widehat{t}) = 1, \quad Q_{3n}(\widehat{\boldsymbol{\beta}}, \widehat{s}, \widehat{t}) = 0, \quad (15)$$

where

$$\begin{aligned} Q_{1n}(\boldsymbol{\beta}, s, t) &= \frac{1}{n} \sum_{i=1}^n \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)} g_i(\boldsymbol{\beta}), \\ Q_{2n}(\boldsymbol{\beta}, s, t) &= \frac{1}{n} \sum_{i=1}^n \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)}, \\ Q_{3n}(\boldsymbol{\beta}, s, t) &= \frac{1}{n(1+\lambda)^2} \\ &\quad \times \sum_{i=1}^n \left\{ 1 + s + t^T g_i(\boldsymbol{\beta}) \right\}^{-1/(1+\lambda)} \\ &\quad \times \left(\frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^T t. \end{aligned} \quad (16)$$

Theorem 2. Suppose that the conditions of Theorem 1 hold, and further assume that $\partial^2 g(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ is continuous about $\boldsymbol{\beta}$ in a neighborhood of the true value $\boldsymbol{\beta}_0$. Then, as $n \rightarrow \infty$, one has

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{L} N(0, \Omega), \quad (17)$$

where “ \xrightarrow{L} ” denotes the convergence in distribution and

$$\Omega = \left[E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right)^T \Sigma^{-1}(\boldsymbol{\beta}_0) E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right) \right]^{-1}. \quad (18)$$

Remark 3. When $\lambda \rightarrow 0$, the empirical log-likelihood ratio statistic is the special case of the empirical Cressie-Read test statistic $CR_\lambda(\boldsymbol{\beta})$.

Remark 4. The proposed method provides a way to find efficient estimates for marginal longitudinal GLMs when the within-subject correlation is considered. It is known that we can construct the confidence regions of $\boldsymbol{\beta}_0$ using Theorem 2. The asymptotic variance Ω of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ does not depend on any λ and can be consistently estimated by

$$\begin{aligned} &\left[\left\{ \sum_{i=1}^n \widehat{p}_i \frac{\partial g_i(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right\}^T \left\{ \sum_{i=1}^n \widehat{p}_i g_i(\widehat{\boldsymbol{\beta}}) g_i^T(\widehat{\boldsymbol{\beta}}) \right\}^{-1} \right. \\ &\quad \left. \times \left\{ \sum_{i=1}^n \widehat{p}_i \frac{\partial g_i(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right\} \right]^{-1} \end{aligned} \quad (19)$$

or by the same expression with the \widehat{p}_i 's, replaced by $1/n$.

Theorem 5. Suppose that the conditions of Theorem 2 hold; then, under the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $CR_\lambda(\boldsymbol{\beta}_0)$ is asymptotically a Chi-squared distribution with r degrees of freedom as $n \rightarrow \infty$.

Theorem 5 allows us to use the empirical Cressie-Read test statistic for testing or obtaining the confidence regions for the parameter $\boldsymbol{\beta}_0$. For any $0 < \alpha < 1$, the confidence region of $\boldsymbol{\beta}_0$ with asymptotic coverage $1 - \alpha$ can be determined by

$$P(CR_\lambda(\boldsymbol{\beta}_0) \leq \chi_r^2) = 1 - \alpha. \quad (20)$$

We can construct a goodness-of-fit statistic to test the model assumption:

$$H_0 : E\{g_i(\boldsymbol{\beta})\} = 0. \quad (21)$$

We call $CR_\lambda(\boldsymbol{\beta}_0) - CR_\lambda(\widehat{\boldsymbol{\beta}})$ the empirical Cressie-Read likelihood ratio statistic and call $CR_\lambda(\widehat{\boldsymbol{\beta}})$ the model test statistic.

Theorem 6. Suppose that the conditions of Theorem 2 hold and $g_i(\boldsymbol{\beta})$ has dimension $r = kp$ and $\boldsymbol{\beta}$ has dimension p with $p < r$. Then, under the model assumption (21), $CR_\lambda(\widehat{\boldsymbol{\beta}})$ is asymptotically a Chi-squared distribution with $r - p$ degrees of freedom as $n \rightarrow \infty$; under the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $CR_\lambda(\boldsymbol{\beta}_0) - CR_\lambda(\widehat{\boldsymbol{\beta}})$ is asymptotically a Chi-squared distribution with p degrees of freedom as $n \rightarrow \infty$.

TABLE 1: The average coverage probabilities (ACP) and the average lengths (ALEN) of the confidence intervals for β_1 , β_2 , and β_3 when the nominal level is 0.95 and the true correlation structure is AR(1) structure.

	n	β	EL		ET		GEE		QIF	
			ACP	ALEN	ACP	ALEN	ACP	ALEN	ACP	ALEN
CS	60	β_1	0.9120	0.4647	0.9060	0.4741	0.9000	0.4658	0.9120	0.4879
		β_2	0.9140	0.6375	0.8820	0.6539	0.8400	0.5561	0.8700	0.6642
		β_3	0.9280	0.4520	0.9150	0.4600	0.9100	0.4565	0.8960	0.4378
	100	β_1	0.9160	0.4646	0.9120	0.4713	0.9100	0.4625	0.9120	0.4742
		β_2	0.9180	0.6396	0.9080	0.6512	0.8940	0.6152	0.8740	0.6569
		β_3	0.9240	0.4518	0.9180	0.4577	0.9100	0.4432	0.9040	0.4432
AR(1)	60	β_1	0.9220	0.3608	0.9180	0.3650	0.9200	0.3663	0.9480	0.3848
		β_2	0.9240	0.4959	0.9120	0.5033	0.8760	0.4288	0.9020	0.5149
		β_3	0.9320	0.3513	0.9280	0.3550	0.9240	0.3599	0.9280	0.3585
	100	β_1	0.9300	0.3613	0.9200	0.3641	0.9220	0.3541	0.9460	0.3716
		β_2	0.9360	0.4964	0.9320	0.5016	0.9160	0.4857	0.9140	0.5128
		β_3	0.9420	0.3514	0.9380	0.3541	0.9240	0.3597	0.9340	0.3537

4. Numerical Studies

4.1. Simulation Studies. In this subsection, we report the simulation study to illustrate the finite sample properties of the proposed ECR test statistic. For simplicity, we only compute the empirical likelihood (EL, $\lambda = 0$) and the empirical exponential tilting (ET, $\lambda \rightarrow -1$) and compare the proposed methods with the GEE and QIF methods. Throughout the simulation study, each dataset comprised $n = 60$ and 100 subjects and $m = 5$ observations per subject over time. For each case, we repeat the experiment 500 times.

Consider the following logistic regression model. The response variable y_{ij} is binary and its marginal expectation given \mathbf{x}_{ij} is

$$\text{logit}(\mu_{ij}) = \sum_{k=1}^3 x_{ij,k} \beta_{0k}, \quad i = 1, \dots, n, \quad j = 1, \dots, 5, \quad (22)$$

where $\beta_0 = [1, 2, -0.8]^T$ and the covariate $\mathbf{x}_{ij} = (x_{ij,1}, \dots, x_{ij,3})^T$ has a multivariate normal distribution with mean zero, marginal variance 1, and an AR(1) correlation matrix with autocorrelation coefficient 0.7. The binary response vector for each cluster has mean specified by (22) and an AR(1) correlation structure with an autocorrelation coefficient $\alpha = 0.7$. Table 1 reports the simulation results for the average coverage probabilities and the average lengths of the confidence intervals for β_1 , β_2 , and β_3 when the nominal level is 0.95 and the true correlation structure is AR(1) structure.

From Table 1, it is easy to see that the empirical likelihood method performs much better in terms of coverage accuracies of the confidence intervals even when the working correlation structure is misspecified. And when the working correlation structure is correctly specified, the performances of all the methods are usually slightly better. Table 1 also shows that the average coverage probabilities obtained by all methods tend to the nominal level 0.95 and the average lengths decrease as n increases.

When sample size is 60 and the true correlation structure is AR(1), the histograms and the QQ plots of the 500 maximum empirical likelihood estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ based

on the empirical likelihood method are plotted in Figure 1 and the histograms and the QQ plots of the 500 maximum empirical exponential tilting estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ based on the empirical exponential tilting method are plotted in Figure 2 under the misspecified and correct correlation structures, respectively. Figures 1 and 2 show that empirically these estimators are asymptotically normal even when the working correlation structure is misspecified.

4.2. Application to Real Data. To examine the performance of the proposed method, we analyze real data [35, 36] from a six-week frequent magnetic resonance imaging (MRI) substudy of the Betaseron clinical trial conducted at the University of British Columbia in relapsing-remitting multiple sclerosis involving 52 patients. The real data concerns a longitudinal clinical trial to assess the effects of neutralizing antibodies on interferon beta-1b (IFNB) in relapsing-remitting multiple sclerosis (MS), which is a disease that destroys the myelin sheath that surrounds the nerves. All patients were randomized into three treatment groups with allocation of 17 patients being treated by placebo, 17 by low dose, and 16 by high dose. This dataset has been studied by Song [37] and Li et al. [11]. There exist the missing values in this dataset; for convenience, we only analyze the balanced longitudinal data which contain 39 patients.

For the analysis of this data, the binary response variable is exacerbation, which refers to whether an exacerbation began since the previous MRI scan, and is 1 for yes and 0 for no. Several baseline covariates are included in the model. They are treatment (trt), time (T) in weeks, squared time (T^2), and duration of disease (dur) in years. Here trt is treated as an ordinal covariate with scale 0, 1, 2 representing zero (placebo), low, and high dosage of the drug treatment. We consider the following marginal logistic model for the data:

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \text{trt}_i + \beta_2 T_j + \beta_3 T_j^2 + \beta_4 \text{dur}_j, \quad (23)$$

where μ_{ij} is the probability of exacerbation at visit j for subject i . Two correlation structures (exchangeable (CS) and AR(1))

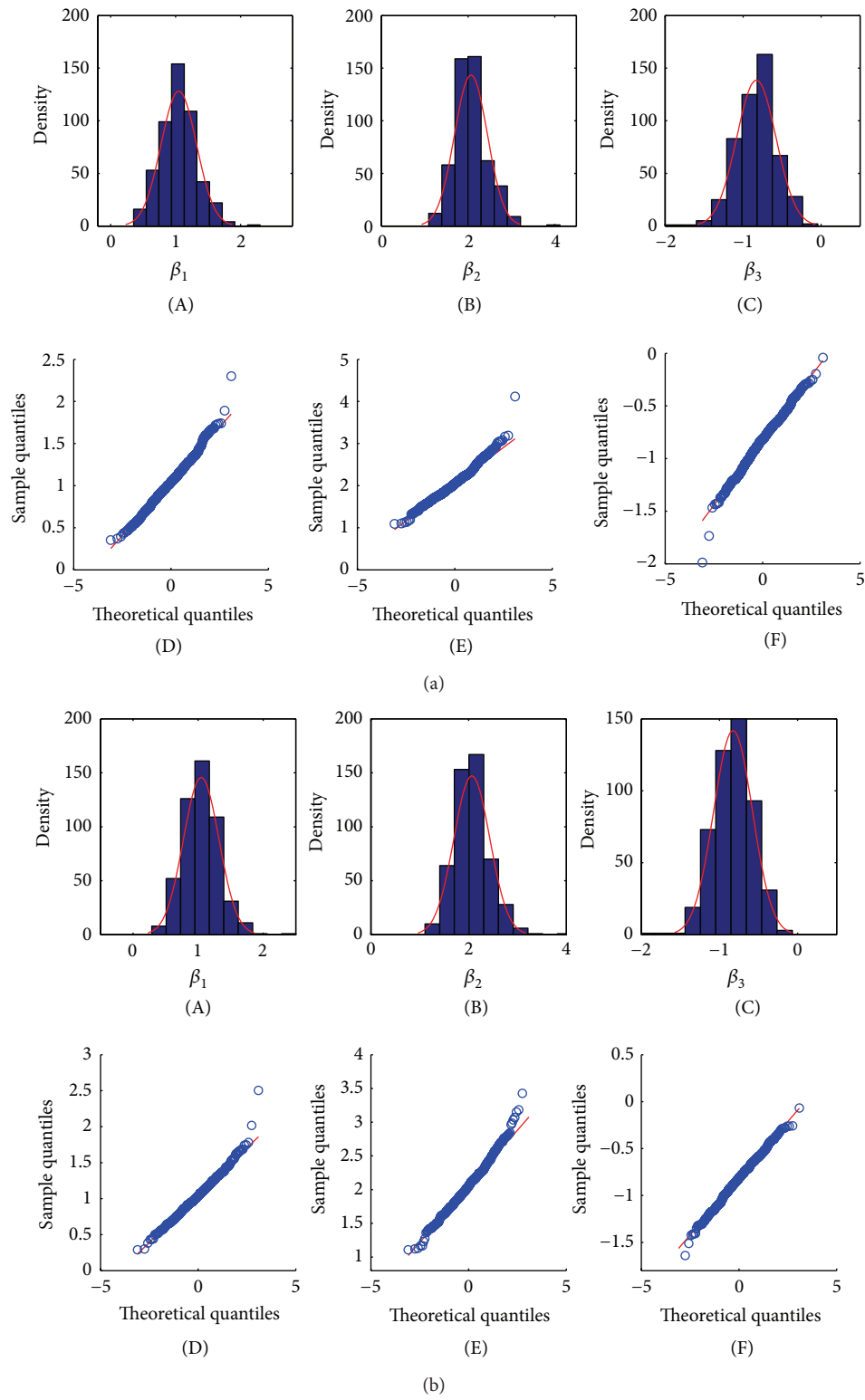


FIGURE 1: The histograms and the QQ plots for the maximum empirical likelihood estimators based on the empirical likelihood method for $n = 60$. When the true correlation structure is AR(1), the left plots are obtained by using the misspecified CS working correlation structure and the right plots are obtained by using the true AR(1) working correlation structure.

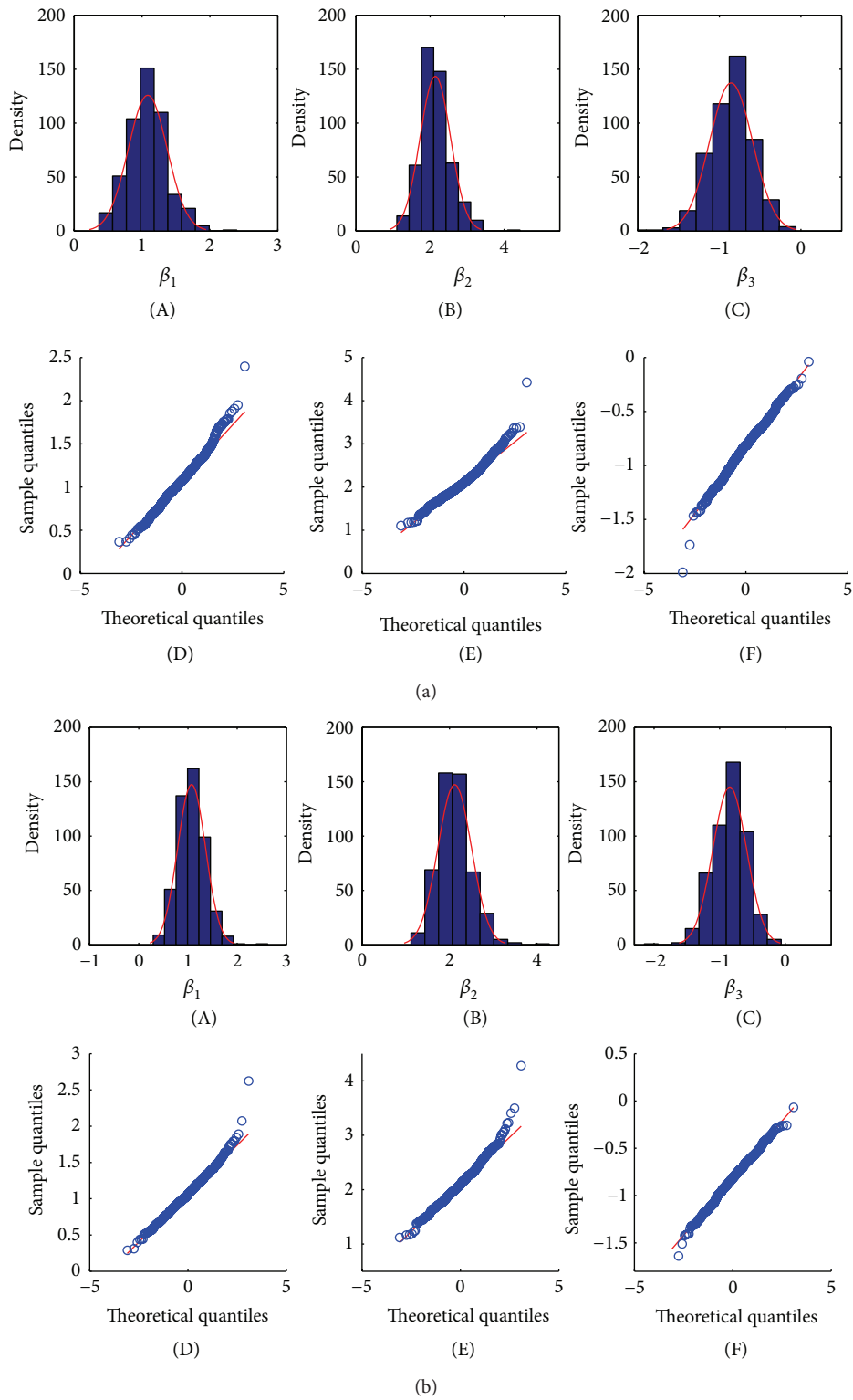


FIGURE 2: The histograms and the QQ plots for the maximum empirical exponential tilting estimators based on the empirical exponential tilting method for $n = 60$. When the true correlation structure is AR(1), the left plots are obtained by using the misspecified CS working correlation structure and the right plots are obtained by using the true AR(1) working correlation structure.

TABLE 2: The parameter estimators (the corresponding standard errors) for CS and AR(1) correlation structures.

		EL	ET	GEE	QIF
CS	$\hat{\beta}_0$	-0.3834 (0.3820)	-0.3051 (0.3872)	-0.4030 (0.4096)	-0.5039 (0.4005)
	$\hat{\beta}_1$	-0.0321 (0.1239)	-0.0454 (0.1259)	-0.0809 (0.1510)	-0.0667 (0.1206)
	$\hat{\beta}_2$	-0.0344 (0.0141)	-0.0349 (0.0143)	-0.0301 (0.0146)	-0.0296 (0.0146)
	$\hat{\beta}_3$	0.00031 (0.00013)	0.00031 (0.00013)	0.00027 (0.00013)	0.00027 (0.00010)
	$\hat{\beta}_4$	-0.0574 (0.0189)	-0.0700 (0.0200)	-0.0617 (0.0227)	-0.0530 (0.0255)
AR(1)	$\hat{\beta}_0$	-0.4212 (0.4016)	-0.4382 (0.4041)	-0.4009 (0.3682)	-0.5043 (0.4145)
	$\hat{\beta}_1$	-0.0590 (0.1236)	-0.0599 (0.1242)	-0.0767 (0.1195)	-0.0711 (0.1506)
	$\hat{\beta}_2$	-0.0295 (0.0146)	-0.0278 (0.0147)	-0.0309 (0.0138)	-0.0290 (0.0089)
	$\hat{\beta}_3$	0.00026 (0.00013)	0.00025 (0.00013)	0.00028 (0.00012)	0.00027 (0.00008)
	$\hat{\beta}_4$	-0.0620 (0.0206)	-0.0645 (0.0208)	-0.0606 (0.0179)	-0.0558 (0.0249)

TABLE 3: 95% confidence intervals (CI) and the lengths (LEN) of the confidence intervals for CS and AR(1) correlation structures.

		EL	ET	GEE	QIF	
CS	β_0	CI	[-1.1321, 0.3654]	[-1.0640, 0.4539]	[-1.2058, 0.3998]	[-1.1504, 0.3596]
		LEN	1.4975	1.5179	1.6056	1.5100
	β_1	CI	[-0.2749, 0.2107]	[-0.2922, 0.2014]	[-0.3768, 0.2150]	[-0.3680, 0.2065]
		LEN	0.4856	0.4936	0.5918	0.5745
	β_2	CI	[-0.0620, -0.0068]	[-0.0629, -0.0069]	[-0.0587, -0.0016]	[-0.0604, -0.0039]
		LEN	0.0552	0.0560	0.0571	0.0565
	β_3	CI	[0.0001, 0.0006]	[0.0001, 0.0006]	[0.00002, 0.00052]	[0.00008, 0.00056]
		LEN	0.0005	0.0005	0.0005	0.00048
	β_4	CI	[-0.0945, -0.0203]	[-0.1092, -0.0308]	[-0.1061, -0.0172]	[-0.1030, -0.0030]
		LEN	0.742	0.0784	0.0889	0.09995
AR(1)	β_0	CI	[-1.2084, 0.3659]	[-1.2302, 0.3537]	[-1.1226, 0.3208]	[-1.2268, 0.2760]
		LEN	1.5743	1.5839	1.4434	1.5028
	β_1	CI	[-0.3012, 0.1832]	[-0.3033, 0.1834]	[-0.3109, 0.1576]	[-0.3663, 0.2240]
		LEN	0.4844	0.4867	0.4685	0.5903
	β_2	CI	[-0.0580, -0.0009]	[-0.0566, 0.0010]	[-0.0580, -0.0037]	[-0.0465, -0.0115]
		LEN	0.0571	0.0576	0.0543	0.0350
	β_3	CI	[0, 0.0005]	[0, 0.0005]	[0.00004, 0.00052]	[0.00011, 0.00052]
		LEN	0.0005	0.0005	0.00048	0.00041
	β_4	CI	[-0.1023, -0.0217]	[-0.1053, -0.0238]	[-0.0957, -0.0255]	[-0.1047, -0.0070]
		LEN	0.0806	0.0815	0.0702	0.0977

are considered in this analysis. Table 2 reports the coefficient estimators and the corresponding standard errors for CS and AR(1) correlation structures, respectively. Table 3 reports the 95% confidence intervals and the lengths of the confidence intervals for CS and AR(1) correlation structures.

From Tables 2 and 3, we can see that the parameter estimators are very similar for all four methods under the CS and AR(1) working correlation structures. For the CS working correlation structure, the empirical likelihood and the empirical exponential tilting methods have smaller standard errors and smaller interval lengths than the GEE and QIF methods. However, the GEE method has the smaller standard errors and the smaller interval lengths under the AR(1) working correlation structure. Similar to the results in Song [37], we also find that the baseline disease severity measured as the duration of disease before the trial is an important explanatory variable associated with the risk of exacerbation,

and we also do not find strong evidence that the drug treatment is effective in reducing the risk of exacerbation. Therefore, all the methods have comparable performance, and these findings are close to the existing analysis in Song [37].

5. Proofs of the Theorems

In this section, we present rigorous proofs of our results stated in Section 3.

Proof of Theorem 1. Assume that $\lambda \neq -1$; G defined by (10) takes derivatives with respect to p_i ; we have

$$\frac{\partial G}{\partial p_i} = -\frac{2}{(1 + \lambda)n^\lambda} p_i^{-(1+\lambda)} + \kappa_1 + n\kappa_2^T g_i(\boldsymbol{\beta}) = 0, \quad (24)$$

which implies that

$$p_i^{-(1+\lambda)} = \frac{(1+\lambda)n^\lambda(\kappa_1 + n\kappa_2^T g_i(\boldsymbol{\beta}))}{2}. \quad (25)$$

Multiplying p_i at both sides of (25) and taking sum and noting that $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i g_i(\boldsymbol{\beta}) = 0$, we can derive that

$$\sum_{i=1}^n p_i^{-\lambda} = \frac{(1+\lambda)n^\lambda \kappa_1}{2}. \quad (26)$$

Substituting (26) into (25), we have

$$p_i^{-(1+\lambda)} = \sum_{i=1}^n p_i^{-\lambda} + \frac{(1+\lambda)n^{1+\lambda}\kappa_2^T g_i(\boldsymbol{\beta})}{2}. \quad (27)$$

By Theorem 1 in Baggerly [17], there exists a constant $c > 0$ such that

$$\frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left[(np_i)^{-\lambda} - 1 \right] \leq c, \quad (28)$$

which implies that $\sum_{i=1}^n p_i^{-\lambda} = n^{\lambda+1} + \eta$, where $\eta = O_P(n^\lambda)$. Substituting this expression into (27), we have

$$p_i^{-(1+\lambda)} = n^{1+\lambda} + \eta + \frac{(1+\lambda)n^{1+\lambda}\kappa_2^T g_i(\boldsymbol{\beta})}{2}. \quad (29)$$

Therefore, we can obtain that

$$p_i = \frac{1}{n} \left[1 + s + t^T g_i(\boldsymbol{\beta}) \right]^{-1/(1+\lambda)}, \quad (30)$$

where $s = \eta/n^{1+\lambda} = O_P(n^{-1})$ and $t = ((1+\lambda)/2)\kappa_2$. The next step is to bound the convergence rate of t . Let $t = \rho t_0$, where $\|t_0\| = 1$. Invoking Lemma 3.2 in Künsch [38] and $E(g_i(\boldsymbol{\beta})) = 0$, it is easy to show that $t_0^T g_i(\boldsymbol{\beta}) = O_P(n^{1/2})$. Applying the Taylor expansion to $\sum_{i=1}^n p_i g_i(\boldsymbol{\beta}) = 0$, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left[g_i(\boldsymbol{\beta}) - \frac{1}{1+\lambda} (s g_i(\boldsymbol{\beta}) + \rho g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) t_0) + \zeta_n \right] \\ & = 0, \end{aligned} \quad (31)$$

where $\zeta_n = \zeta_n(g_i(\boldsymbol{\beta}), s, \rho)$. Assume that $\rho = O_P(n^{-\nu})$, $\nu > 0$. By the central limit theorem, we have $(1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta}) = O_P(n^{-1/2})$ and $(1/n) \sum_{i=1}^n s g_i(\boldsymbol{\beta}) = O_P(n^{-3/2})$. Let $\Sigma(\boldsymbol{\beta}) = \text{Var}\{g_i(\boldsymbol{\beta})\}$; it is easy to show that

$$\widehat{\Sigma} \triangleq \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) \longrightarrow \Sigma(\boldsymbol{\beta}), \quad \text{a.s.} \quad (32)$$

If $\nu > 1/2$ (or $\nu < 1/2$), then the dominant term in the right-hand side of (31) is $(1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta}) = O_P(n^{-1/2})$, while $\rho \widehat{\Sigma} t_0 = O_P(n^{-\nu})$, which implies that the sum cannot be zero unless $\nu = 1/2$. Therefore, we obtain that $\rho = O_P(n^{-1/2})$. By (31), we have

$$\rho t_0 = (1+\lambda) \widehat{\Sigma}^{-1} \left(\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right) + \zeta_n, \quad \zeta_n = o_P(n^{-1/2}). \quad (33)$$

For convenience, let $\bar{g}_n(\boldsymbol{\beta}) \triangleq (1/n) \sum_{i=1}^n g_i(\boldsymbol{\beta})$. We now find an approximation for s in terms of $\rho t_0^T g_i(\boldsymbol{\beta}) = O_P(1)$. By $\sum_{i=1}^n p_i = 1$ and applying the Taylor expansion, we have

$$\frac{1}{n} \sum_{i=1}^n \left[1 - \frac{d_n}{1+\lambda} + \frac{(2+\lambda)d_n^2}{2(1+\lambda)^2} + O_P(n^{-1/2}\zeta_n) \right] = 1, \quad (34)$$

where $d_n = s + \rho t_0^T g_i(\boldsymbol{\beta})$ and ζ_n is defined by (33). Rearranging (34) and ignoring terms of order $O_P(s^2)$, we have

$$\begin{aligned} & s + \rho t_0^T \bar{g}_n(\boldsymbol{\beta}) \\ & = \frac{2+\lambda}{2(1+\lambda)} \left[2s \rho t_0^T \bar{g}_n(\boldsymbol{\beta}) \right. \\ & \quad \left. + \frac{1}{n} \rho^2 t_0^T \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) t_0 \right] \\ & \quad + O_P(n^{-1/2}\zeta_n). \end{aligned} \quad (35)$$

That is,

$$s = \frac{2+\lambda}{2(1+\lambda)} \rho^2 t_0^T \widehat{\Sigma} t_0 - \rho t_0^T \bar{g}_n(\boldsymbol{\beta}) + O_P(n^{-1/2}\zeta_n). \quad (36)$$

From (33) and (36), we have

$$s = \frac{\lambda}{2} \rho t_0^T \bar{g}_n(\boldsymbol{\beta}) + O_P(n^{-1/2}\zeta_n). \quad (37)$$

By (33) and (37) and applying the Taylor expansion for $\text{CR}_\lambda(\boldsymbol{\beta})$, it is easy to show that

$$\begin{aligned} \text{CR}_\lambda(\boldsymbol{\beta}) &= \frac{2}{\lambda(1+\lambda)} \sum_{i=1}^n \left[(1 + s + \rho t_0^T g_i(\boldsymbol{\beta}))^{\lambda/(1+\lambda)} - 1 \right] \\ &= \frac{2}{(1+\lambda)^2} \sum_{i=1}^n \left[s + \rho t_0^T g_i(\boldsymbol{\beta}) - \frac{1}{2(1+\lambda)} \right. \\ & \quad \left. \times \rho^2 t_0^T g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) t_0 + o_P(n^{-1}) \right] \\ &= \frac{n}{(1+\lambda)^2} \left[2s + 2\rho t_0^T \bar{g}_n(\boldsymbol{\beta}) - \frac{1}{1+\lambda} \rho^2 t_0^T \widehat{\Sigma} t_0 \right. \\ & \quad \left. + o_P(n^{-1}) \right] \\ &= \frac{n}{(1+\lambda)^2} \left[\lambda \rho t_0^T \bar{g}_n(\boldsymbol{\beta}) \right. \\ & \quad + 2(1+\lambda) \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) \\ & \quad + 2O_P(n^{-1/2})\zeta_n \\ & \quad - (1+\lambda) \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) \\ & \quad - 2\bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \zeta_n - \zeta_n^T \widehat{\Sigma}^{-1} \zeta_n \\ & \quad \left. + o_P(n^{-1}) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{n}{(1+\lambda)^2} \left[\lambda(1+\lambda) \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) \right. \\
&\quad + 2(1+\lambda) \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) \\
&\quad \left. - (1+\lambda) \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) \right] + o_P(1) \\
&= n \bar{g}_n^T(\boldsymbol{\beta}) \widehat{\Sigma}^{-1} \bar{g}_n(\boldsymbol{\beta}) + o_P(1). \tag{38}
\end{aligned}$$

Similar to the proof of Lemma 1 in Qin and Lawless [20], let $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + un^{-1/3}$, for $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = n^{-1/3}\}$, where $\|u\| = 1$. We first give a lower bound for $CR_\lambda(\boldsymbol{\beta})$ on the surface of the ball. When $E\|g_i(\boldsymbol{\beta})\|^3 < \infty$ and $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq n^{-1/3}$, we can obtain that

$$\begin{aligned}
t(\boldsymbol{\beta}) &= (1+\lambda) \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) \right]^{-1} \\
&\quad \times \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right] + o(n^{-1/3}) \tag{39} \\
&= O(n^{-1/3}), \quad \text{a.s.}
\end{aligned}$$

holds uniformly for $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = n^{-1/3}\}$. From (32), (38), and (39), we have

$$\begin{aligned}
CR_\lambda(\boldsymbol{\beta}) &= n \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right]^T \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) \right]^{-1} \\
&\quad \times \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) \right] + o(n^{-1/3}) \\
&= n \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}_0) + \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} un^{-1/3} \right]^T \\
&\quad \times \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}) \right]^{-1} \\
&\quad \times \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}_0) + \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} un^{-1/3} \right] \\
&\quad + o(n^{-1/3}) \tag{40} \\
&= n \left[O(n^{-1/2}(\log \log n)^{1/2}) \right. \\
&\quad \left. + E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right) un^{-1/3} \right]^T \Sigma^{-1}(\boldsymbol{\beta}_0) \\
&\quad \times \left[O(n^{-1/2}(\log \log n)^{1/2}) \right. \\
&\quad \left. + E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right) un^{-1/3} \right] + o(n^{-1/3}) \\
&\geq (\phi - \epsilon) n^{1/3}, \quad \text{a.s.},
\end{aligned}$$

where $\phi - \epsilon > 0$ and ϕ is the smallest eigenvalue of

$$E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right)^T \Sigma^{-1}(\boldsymbol{\beta}_0) E \left(\frac{\partial g_1(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \right). \tag{41}$$

Similarly,

$$\begin{aligned}
CR_\lambda(\boldsymbol{\beta}_0) &= n \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}_0) \right]^T \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}_0) g_i^T(\boldsymbol{\beta}_0) \right]^{-1} \\
&\quad \times \left[\frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\beta}_0) \right] + o(1) \tag{42} \\
&= O(\log \log n), \quad \text{a.s.}
\end{aligned}$$

It is known that, when $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq n^{-1/3}\}$, $CR_\lambda(\boldsymbol{\beta})$ is a continuous function about $\boldsymbol{\beta}$ and has a minimum value in the interior of this ball. In addition, $\hat{\boldsymbol{\beta}}$ satisfies

$$\begin{aligned}
\frac{\partial CR_\lambda(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= \frac{2}{(1+\lambda)^2} \sum_{i=1}^n [1 + s + t^T g_i(\boldsymbol{\beta})]^{-1/(1+\lambda)} \\
&\quad \times \left(\frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^T t \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0. \tag{43}
\end{aligned}$$

□

Proof of Theorem 2. We take derivatives for Q_{1n} , Q_{2n} , and Q_{3n} with respect to $\boldsymbol{\beta}$, s , and t^T ; we have

$$\begin{aligned}
\frac{\partial Q_{1n}(\boldsymbol{\beta}, 0, 0)}{\partial \boldsymbol{\beta}} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \\
\frac{\partial Q_{1n}(\boldsymbol{\beta}, 0, 0)}{\partial s} &= -\frac{1}{n(1+\lambda)} \sum_{i=1}^n g_i(\boldsymbol{\beta}), \\
\frac{\partial Q_{1n}(\boldsymbol{\beta}, 0, 0)}{\partial t^T} &= -\frac{1}{n(1+\lambda)} \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i^T(\boldsymbol{\beta}); \\
\frac{\partial Q_{2n}(\boldsymbol{\beta}, 0, 0)}{\partial \boldsymbol{\beta}} &= 0, \\
\frac{\partial Q_{2n}(\boldsymbol{\beta}, 0, 0)}{\partial s} &= -\frac{1}{1+\lambda}, \\
\frac{\partial Q_{2n}(\boldsymbol{\beta}, 0, 0)}{\partial t^T} &= -\frac{1}{n(1+\lambda)} \sum_{i=1}^n g_i^T(\boldsymbol{\beta}); \\
\frac{\partial Q_{3n}(\boldsymbol{\beta}, 0, 0)}{\partial \boldsymbol{\beta}} &= 0, \quad \frac{\partial Q_{3n}(\boldsymbol{\beta}, 0, 0)}{\partial s} = 0, \\
\frac{\partial Q_{3n}(\boldsymbol{\beta}, 0, 0)}{\partial t^T} &= \frac{1}{n(1+\lambda)^2} \sum_{i=1}^n \left(\frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^T. \tag{44}
\end{aligned}$$

By the conditions and Theorem 1 and the Taylor expanding of $Q_{1n}(\hat{\beta}, \hat{s}, \hat{t})$, $Q_{2n}(\hat{\beta}, \hat{s}, \hat{t})$, and $Q_{3n}(\hat{\beta}, \hat{s}, \hat{t})$ at $(\beta_0, 0, 0)$, we have

$$\begin{aligned} 0 &= Q_{1n}(\hat{\beta}, \hat{s}, \hat{t}) \\ &= Q_{1n}(\beta_0, 0, 0) + \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial \beta} (\hat{\beta} - \beta_0) \\ &\quad + \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial s} (\hat{s} - 0) \\ &\quad + \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial t^T} (\hat{t} - 0) + o_P(\delta_n), \end{aligned}$$

$$\begin{aligned} 1 &= Q_{2n}(\hat{\beta}, \hat{s}, \hat{t}) \\ &= Q_{2n}(\beta_0, 0, 0) + \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial \beta} (\hat{\beta} - \beta_0) \\ &\quad + \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial s} (\hat{s} - 0) \\ &\quad + \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial t^T} (\hat{t} - 0) + o_P(\delta_n), \end{aligned}$$

$$\begin{aligned} 0 &= Q_{3n}(\hat{\beta}, \hat{s}, \hat{t}) \\ &= Q_{3n}(\beta_0, 0, 0) + \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial \beta} (\hat{\beta} - \beta_0) \\ &\quad + \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial s} (\hat{s} - 0) \\ &\quad + \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial t^T} (\hat{t} - 0) + o_P(\delta_n), \end{aligned}$$

(45)

where $\delta_n = \|\hat{\beta} - \beta_0\| + \|\hat{s}\| + \|\hat{t}\|$. We further have

$$\begin{aligned} &\begin{pmatrix} -Q_{1n}(\beta_0, 0, 0) + o_P(\delta_n) \\ 1 - Q_{2n}(\beta_0, 0, 0) + o_P(\delta_n) \\ -Q_{3n}(\beta_0, 0, 0) + o_P(\delta_n) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial \beta} \\ \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial \beta} \\ \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial \beta} \end{pmatrix} \\ &\quad \times \begin{pmatrix} \hat{t} \\ \hat{s} \\ \hat{\beta} - \beta_0 \end{pmatrix}. \end{aligned}$$

(46)

By the above expression, we can obtain that

$$\begin{pmatrix} \hat{t} \\ \hat{s} \\ \hat{\beta} - \beta_0 \end{pmatrix} = S_n^{-1} \begin{pmatrix} -Q_{1n}(\beta_0, 0, 0) + o_P(\delta_n) \\ o_P(\delta_n) \\ o_P(\delta_n) \end{pmatrix}, \quad (47)$$

where

$$\begin{aligned} S_n &\triangleq \begin{pmatrix} \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{1n}(\beta_0, 0, 0)}{\partial \beta} \\ \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{2n}(\beta_0, 0, 0)}{\partial \beta} \\ \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial t^T} & \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial s} & \frac{\partial Q_{3n}(\beta_0, 0, 0)}{\partial \beta} \end{pmatrix} \\ &\rightarrow S, \quad \text{a.s.}, \\ S &= \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & 0 \\ S_{31} & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{1+\lambda} \Sigma(\beta_0) & -\frac{1}{1+\lambda} E(g_1(\beta_0)) & E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right) \\ -\frac{1}{1+\lambda} E(g_1^T(\beta_0)) & -\frac{1}{1+\lambda} & 0 \\ \frac{1}{(1+\lambda)^2} E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right)^T & 0 & 0 \end{pmatrix}. \end{aligned} \quad (48)$$

By the above expression and $Q_{1n}(\beta_0, 0, 0) = (1/n) \sum_{i=1}^n g_i(\beta_0) = O_P(n^{-1/2})$, we can derive that $\delta_n = O_P(n^{-1/2})$. Note that $E(g_i(\beta_0)) = 0$, and $\Sigma(\beta_0)$ is the positive definite matrix; it is easy to show that

$$S^{-1} = \begin{pmatrix} S_{11}^{-1} + S_{11}^{-1} [S_{13} S_{33.1}^{-1} S_{31}] S_{11}^{-1} & \mathbf{0} & -S_{11}^{-1} S_{13} S_{33.1}^{-1} \\ \mathbf{0} & S_{22}^{-1} & 0 \\ -S_{33.1}^{-1} S_{31} S_{11}^{-1} & 0 & S_{33.1}^{-1} \end{pmatrix}, \quad (49)$$

where $S_{33.1} = -S_{31} S_{11}^{-1} S_{13}$. Substituting (49) into (47), we have

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta_0) &= S_{33.1}^{-1} S_{31} S_{11}^{-1} \sqrt{n} Q_{1n}(\beta_0, 0, 0) + o_P(1), \\ \sqrt{n}(\hat{t} - 0) &= -\{S_{11}^{-1} + S_{11}^{-1} [S_{13} S_{33.1}^{-1} S_{31}] S_{11}^{-1}\} \\ &\quad \times \sqrt{n} Q_{1n}(\beta_0, 0, 0) + o_P(1). \end{aligned} \quad (50)$$

Note that $\sqrt{n} Q_{1n}(\beta_0, 0, 0) \xrightarrow{L} N(0, \Sigma(\beta_0))$; we can derive that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, \Omega), \quad \sqrt{n}(\hat{t} - 0) \xrightarrow{L} N(0, \Phi), \quad (51)$$

where

$$\begin{aligned} \Omega &= \left[E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right)^T \Sigma^{-1}(\beta_0) E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right) \right]^{-1}, \\ \Phi &= (1+\lambda) \Sigma^{-1}(\beta_0) \\ &\quad \times \left\{ I - E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right) \Omega E\left(\frac{\partial g_1(\beta_0)}{\partial \beta}\right)^T \Sigma^{-1}(\beta_0) \right\}. \end{aligned} \quad (52)$$

□

Proof of Theorem 5. Since $\sqrt{n}\bar{g}_n(\beta_0) \xrightarrow{L} N(0, \Sigma(\beta_0))$ and $\widehat{\Sigma}(\beta_0) \stackrel{\Delta}{=} (1/n) \sum_{i=1}^n g_i(\beta_0)g_i^T(\beta_0) \rightarrow \Sigma(\beta_0)$, a.s., by (38), it is easy to show that

$$\begin{aligned} CR_\lambda(\beta_0) &= [\sqrt{n}\bar{g}_n(\beta_0)]^T \widehat{\Sigma}^{-1}(\beta_0) \\ &\quad \times [\sqrt{n}\bar{g}_n(\beta_0)] + o_P(1) \xrightarrow{L} \chi_r^2. \end{aligned} \tag{53}$$

Proof of Theorem 6. From (33), (38), and (50) and by the Taylor expansion and some simple calculations, we have

$$\begin{aligned} CR_\lambda(\hat{\beta}) &= n\bar{g}_n^T(\hat{\beta}) \widehat{\Sigma}^{-1}(\hat{\beta}) \bar{g}_n(\hat{\beta}) + o_P(1) \\ &= \frac{n}{1+\lambda} \bar{g}_n^T(\hat{\beta}) \hat{t} + o_P(1) \\ &= \frac{n}{1+\lambda} \left[\bar{g}_n(\beta_0) + \frac{\partial \bar{g}_n(\beta_0)}{\partial \beta} (\hat{\beta} - \beta_0) \right. \\ &\quad \left. + o_P(n^{-1/2}) \right]^T \hat{t} + o_P(1) \\ &= \frac{n}{1+\lambda} \left[Q_{1n}(\beta_0, 0, 0) \right. \\ &\quad \left. + S_{13} S_{33.1}^{-1} S_{31} S_{11}^{-1} Q_{1n}(\beta_0, 0, 0) \right. \\ &\quad \left. + o_P(n^{-1/2}) \right]^T \\ &\quad \times \left[-\left\{ S_{11}^{-1} + S_{11}^{-1} \left[S_{13} S_{33.1}^{-1} S_{31} \right] S_{11}^{-1} \right\} Q_{1n} \right. \\ &\quad \left. \times (\beta_0, 0, 0) + o_P(n^{-1/2}) \right] + o_P(1) \\ &= \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right]^T \\ &\quad \times G_1 \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right] + o_P(1), \end{aligned} \tag{54}$$

since $G_1^2 = G_1$, where $G_1 \stackrel{\Delta}{=} [I - \Sigma^{-1/2}(\beta_0)E(\partial g_1(\beta_0)/\partial \beta)\Omega E(\partial g_1(\beta_0)/\partial \beta)^T \Sigma^{-1/2}(\beta_0)]$. From $\sqrt{n}\Sigma^{-1/2}(\beta_0)Q_{1n}(\beta_0, 0, 0) \xrightarrow{L} N(0, I)$ and

$$\begin{aligned} \text{tr}(G_1) &= \text{tr} \left\{ I - \Sigma^{-1/2}(\beta_0) E \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right) \Omega E \right. \\ &\quad \left. \times \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right)^T \Sigma^{-1/2}(\beta_0) \right\} \\ &= r - \text{tr} \left\{ \Sigma^{-1/2}(\beta_0) E \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right) \Omega E \right. \\ &\quad \left. \times \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right)^T \Sigma^{-1/2}(\beta_0) \right\} \\ &= r - p, \end{aligned} \tag{55}$$

it is easy to show that $CR_\lambda(\hat{\beta}) \xrightarrow{L} \chi_{r-p}^2$.

On the other hand, invoking the same argument, we have

$$\begin{aligned} CR_\lambda(\beta_0) - CR_\lambda(\hat{\beta}) &= n\bar{g}_n^T(\beta_0) \widehat{\Sigma}^{-1}(\beta_0) \bar{g}_n(\beta_0) \\ &\quad - n\bar{g}_n^T(\hat{\beta}) \widehat{\Sigma}^{-1}(\hat{\beta}) \bar{g}_n(\hat{\beta}) + o_P(1) \\ &= nQ_{1n}^T(\beta_0, 0, 0) \Sigma^{-1}(\beta_0) Q_{1n}(\beta_0, 0, 0) \\ &\quad - \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right]^T \\ &\quad \times G_1 \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right] + o_P(1) \\ &= \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right]^T \\ &\quad \times G_2 \left[\sqrt{n}\Sigma^{-1/2}(\beta_0) Q_{1n}(\beta_0, 0, 0) \right] + o_P(1), \end{aligned} \tag{56}$$

since $G_2^2 = G_2$, where $G_2 \stackrel{\Delta}{=} \Sigma^{-1/2}(\beta_0)E(\partial g_1(\beta_0)/\partial \beta)\Omega E(\partial g_1(\beta_0)/\partial \beta)^T \Sigma^{-1/2}(\beta_0)$. Further,

$$\begin{aligned} \text{tr} \left\{ \Sigma^{-1/2}(\beta_0) E \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right) \Omega E \left(\frac{\partial g_1(\beta_0)}{\partial \beta} \right)^T \right. \\ \left. \times \Sigma^{-1/2}(\beta_0) \right\} = p. \end{aligned} \tag{57}$$

Thus, we can obtain that $CR_\lambda(\beta_0) - CR_\lambda(\hat{\beta}) \xrightarrow{L} \chi_p^2$. □

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Junhua Zhang's research was supported by the National Nature Science Foundation of China (11002005) and the Training Programme Foundation for the Beijing Municipal Excellent Talents (2013D005007000005). Ruiqin Tian, Suigen Yang, and Sanying Feng's researches were supported by the National Nature Science Foundation of China (11101014), the Natural Science Foundation of Beijing (1142002), the Science and Technology Project of Beijing Municipal Education Commission (KM201410005010), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20101103120016), and the Doctoral Fund of Innovation of Beijing University of Technology.

References

- [1] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, UK, 2nd edition, 1989.
- [2] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

- [3] S. L. Zeger and K. Y. Liang, "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics*, vol. 42, no. 1, pp. 121–130, 1986.
- [4] A. Qu, B. G. Lindsay, and B. Li, "Improving generalised estimating equations using quadratic inference functions," *Biometrika*, vol. 87, no. 4, pp. 823–836, 2000.
- [5] G. M. Fitzmaurice, "A caveat concerning independence estimating equations with multivariate binary data," *Biometrics*, vol. 51, no. 1, pp. 309–317, 1995.
- [6] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.
- [7] R. M. Balan and I. Schiopu-Kratina, "Asymptotic results with generalized estimating equations for longitudinal data," *The Annals of Statistics*, vol. 33, no. 2, pp. 522–541, 2005.
- [8] A. Qu and R. Li, "Quadratic inference functions for varying-coefficient models with longitudinal data," *Biometrics*, vol. 62, no. 2, pp. 379–391, 2006.
- [9] L. Wang and A. Qu, "Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach," *Journal of the Royal Statistical Society B*, vol. 71, no. 1, pp. 177–190, 2009.
- [10] L. Wang, "GEE analysis of clustered binary data with diverging number of covariates," *The Annals of Statistics*, vol. 39, no. 1, pp. 389–417, 2011.
- [11] G. R. Li, H. Lian, S. Y. Feng, and L. X. Zhu, "Automatic variable selection for longitudinal generalized linear models," *Computational Statistics & Data Analysis*, vol. 61, pp. 174–186, 2013.
- [12] A. B. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.
- [13] A. B. Owen, "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.
- [14] A. B. Owen, "Empirical likelihood for linear models," *The Annals of Statistics*, vol. 19, no. 4, pp. 1725–1747, 1991.
- [15] E. D. Kolaczyk, "Empirical likelihood for generalized linear models," *Statistica Sinica*, vol. 4, no. 1, pp. 199–218, 1994.
- [16] S. X. Chen and H. J. Cui, "An extended empirical likelihood for generalized linear models," *Statistica Sinica*, vol. 13, no. 1, pp. 69–81, 2003.
- [17] K. A. Baggerly, "Empirical likelihood as a goodness-of-fit measure," *Biometrika*, vol. 85, no. 3, pp. 535–547, 1998.
- [18] T. DiCiccio, P. Hall, and J. Romano, "Empirical likelihood is Bartlett-correctable," *The Annals of Statistics*, vol. 19, no. 2, pp. 1053–1061, 1991.
- [19] J. H. Chen and J. Qin, "Empirical likelihood estimation for finite populations and the effective usage of auxiliary information," *Biometrika*, vol. 80, no. 1, pp. 107–116, 1993.
- [20] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *The Annals of Statistics*, vol. 22, no. 1, pp. 300–325, 1994.
- [21] L. X. Zhu and L. G. Xue, "Empirical likelihood confidence regions in a partially linear single-index model," *Journal of the Royal Statistical Society B*, vol. 68, no. 3, pp. 549–570, 2006.
- [22] G. R. Li, L. Lin, and L. X. Zhu, "Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters," *Journal of Multivariate Analysis*, vol. 105, pp. 85–111, 2012.
- [23] A. B. Owen, *Empirical Likelihood*, Chapman & Hall/CRC, New York, NY, USA, 2001.
- [24] L. G. Xue and L. X. Zhu, "Empirical likelihood for a varying coefficient model with longitudinal data," *Journal of the American Statistical Association*, vol. 102, no. 478, pp. 642–654, 2007.
- [25] G. R. Li, P. Tian, and L. G. Xue, "Generalized empirical likelihood inference in semiparametric regression model for longitudinal data," *Acta Mathematica Sinica (English Series)*, vol. 24, no. 12, pp. 2029–2040, 2008.
- [26] G. R. Li, L. X. Zhu, L. G. Xue, and S. Y. Feng, "Empirical likelihood inference in partially linear single-index models for longitudinal data," *Journal of Multivariate Analysis*, vol. 101, no. 3, pp. 718–732, 2010.
- [27] S. Wang, L. Qian, and R. J. Carroll, "Generalized empirical likelihood methods for analyzing longitudinal data," *Biometrika*, vol. 97, no. 1, pp. 79–93, 2010.
- [28] F. Bravo, "Blockwise empirical Cressie-Read test statistics for α -mixing processes," *Statistics & Probability Letters*, vol. 58, no. 3, pp. 319–325, 2002.
- [29] L. P. Hansen, "Large sample properties of generalized method of moments estimators," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.
- [30] A. Qu and P. X.-K. Song, "Assessing robustness of generalised estimating equations and quadratic inference functions," *Biometrika*, vol. 91, no. 2, pp. 447–459, 2004.
- [31] A. Qu and B. G. Lindsay, "Building adaptive estimating equations when inverse of covariance estimation is difficult," *Journal of the Royal Statistical Society B*, vol. 65, no. 1, pp. 127–142, 2003.
- [32] T. R. C. Read and N. A. C. Cressie, *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer, New York, NY, USA, 1991.
- [33] G. W. Imbens, R. H. Spady, and P. Johnson, "Information-theoretic approaches to inference in moment condition models," *Econometrica*, vol. 66, no. 2, pp. 333–357, 1998.
- [34] S. M. Schennach, "Point estimation with exponentially tilted empirical likelihood," *The Annals of Statistics*, vol. 35, no. 2, pp. 634–672, 2007.
- [35] A. J. Petkau, R. A. White, G. C. Ebers et al., "Longitudinal analyses of the effects of neutralizing antibodies on interferon beta-1b in relapsing-remitting multiple sclerosis," *Multiple Sclerosis*, vol. 10, pp. 126–138, 2004.
- [36] A. J. Petkau, "Statistical approaches to assessing the effects of neutralizing antibodies: IFN β -1b in the pivotal trial of relapsing-remitting multiple sclerosis," *Neurology*, vol. 61, no. 9, pp. S35–S37, 2003.
- [37] P. X.-K. Song, *Correlated Data Analysis: Modeling, Analytics and Applications*, Springer, New York, NY, USA, 2007.
- [38] H. R. Künsch, "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, vol. 17, no. 3, pp. 1217–1241, 1989.