

Empirical Likelihood Inference for Generalized Partially Linear Models with Longitudinal Data

Jinghua Zhang^{1,2}, Liugen Xue¹

¹College of Applied Sciences, Beijing University of Technology, Beijing, China

²Department of Information Engineering, Jingdezhen Ceramic Institute, Jiangxi, China

Email: crashdive@126.com, liugenxue@bjut.edu.cn

How to cite this paper: Zhang, J.H. and Xue, L.G. (2020) Empirical Likelihood Inference for Generalized Partially Linear Models with Longitudinal Data. *Open Journal of Statistics*, 10, 188-202.

<https://doi.org/10.4236/ojs.2020.102014>

Received: March 1, 2020

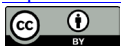
Accepted: March 31, 2020

Published: April 3, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this article, we propose a generalized empirical likelihood inference for the parametric component in semiparametric generalized partially linear models with longitudinal data. Based on the extended score vector, a generalized empirical likelihood ratios function is defined, which integrates the within-cluster correlation meanwhile avoids direct estimating the nuisance parameters in the correlation matrix. We show that the proposed statistics are asymptotically Chi-squared under some suitable conditions, and hence it can be used to construct the confidence region of parameters. In addition, the maximum empirical likelihood estimates of parameters and the corresponding asymptotic normality are obtained. Simulation studies demonstrate the performance of the proposed method.

Keywords

Longitudinal Data, Generalized Partially Linear Models, Empirical Likelihood, Quadratic Inference Function

1. Introduction

An important issue in statistical inference is to construct the confidence region for parameters of interest. The convention method is the normal approximation method which based on the asymptotic normal distribution of parameter estimators. The normal approximation (NA) method requires estimating the limiting variance of regression parameter, which is very complicated in some situation. Besides, the confidence region derived from the NA method is predetermined to be symmetric.

As a nonparametric data-driven technique, the empirical likelihood (EL) approach employs empirical likelihood function without specifically assuming a

distribution for the data, while it can incorporate the side information through constraints, which maximizes the efficiency of the method. Compare with the NA method, EL approach does not involve a plug-in estimation for the limiting variance, and the shapes and the orientation of the confidence region obtained are automatically determined by the data. There has been a lot of literature in empirical likelihood, e.g., [1]-[12].

Longitudinal data often occurs in biomedical research where the repeated measurements from subjects are collected over times, and therefore the responses from same subjects are very likely to be correlated with an unknown structure. The challenge for longitudinal data lies in how to effectively utilize the within-cluster information. The early works in EL for longitudinal data ignored the correlations within subjects, e.g. [7] [8]. Some recent studies incorporate the correlation information by constructing the auxiliary random vector through the generalized estimating equations (GEEs) [9] [10]. The GEEs use a working correlated matrix to carry the correlation information. The working correlated matrix is decided by a small set of nuisance parameters α to avoid the specification of the whole correlation matrix [13]. The advantage of the GEEs is that the estimators of the regression parameter β are always consistent. However, GEEs estimator suffers a great loss in efficiency when the correlation structure is misspecified. The quadratic inference functions (QIFs) approach avoids estimating the nuisance correlation parameters α by assuming that the inverse of the working correlation matrix can be approximated by a linear combination of several known basis matrices, and solve the combined estimation functions by using the principle of the generalized method of moments [14] [15]. The QIFs can also take the within-cluster correlation into account and is more efficient than GEEs when the working correlation is misspecified. The QIFs approach has been applied to many models, including varying coefficient models, partially linear models, single-index models and generalized linear models. The recent related works include [16]-[21]. More recently, [11] [12] proposed generalized empirical likelihood method (GEL), which using a QIFs-based generalized log-empirical likelihood ratio statistics to construct the confidence region for the parameters in generalized linear models (GLMs) with longitudinal data and partially linear models with Longitudinal data.

Generalized partially linear models (GPLMs) can be regarded as a comprise between the GLMs and fully nonparametric models. The choice of a partial linear model is sometimes made to avoid nonparametric specification of high-dimensional covariates, and at other times the model arises naturally due to categorical covariates. In this article, we extend the GEL method to GPLMs with longitudinal data and the B-spline method is adopted to approximate the nonparametric component in the model. Our method incorporates the within-cluster correlation information into the auxiliary random vector. Our proposed method does not require the estimation of the variance of the proposed estimator and is not sensitive to the misspecification of the working correlation structure.

The rest of this article is organized as follows. We propose the QIF-based EL method for GPLMs in Section 2 and present the corresponding asymptotic results in Section 3. Simulation studies are provided in Section 4 and a real data is analysed in Section 5. The details of the proofs are provided in the Appendix.

2. Model and Generalized Empirical Likelihood

2.1. GPLMs with Longitudinal Data

In this article, we consider a longitudinal study with n subjects and m_i observations over time for the i th subject ($i = 1, \dots, n$), for a total of $N = \sum_{i=1}^n m_i$ observations. Each observation consists of a response variable Y_{ij} and the covariate vectors (X_{ij}, U_{ij}) , where $X_{ij} \in R^p$ and U_{ij} is a scalar. We assume that the observations from different subjects are independent, but those from the same subject are dependent. The generalized partially linear model (GPLMs) with longitudinal data take the form

$$\mu_{ij} = E(Y_{ij} | X_{ij}, U_{ij}) = h(X_{ij}^T \beta + \alpha(U_{ij})), \text{var}(Y_{ij} | X_{ij}, U_{ij}) = v(\mu_{ij}). \tag{1}$$

where $\beta = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, $h(\cdot)$ is a known monotonic smooth link function, $\alpha(\cdot)$ is a unknown smooth function and $v(\cdot)$ is a known function with $v(\cdot) > 0$. Without loss of generality, we assume $U \sim U(0,1)$.

Following [22], we replace $\alpha(\cdot)$ by its basis function approximations. More specifically, let $B(u) = (B_1(u), \dots, B_L(u))^T$ be the B-spline basis functions with the order of M , where $L = K + M + 1$, and K is the number of interior knots. We use the B-spline basis functions because they often provide good approximations with a small number of knots. Besides, the B-spline basis functions have bounded support and are numerically stable. The spline approach also treats a non-parametric function as a linear function with the basis functions being the pseudo-design variables, thus any computational algorithm developed for the generalized linear models can be used for the generalized partially linear models.

Suppose $\alpha(u)$ can be approximated by $\alpha(u) \approx B(u)^T \gamma$, where $\gamma = (\gamma_1, \dots, \gamma_L)^T$ is a $L \times 1$ vector of unknown regression coefficients. Then our regression model (1) becomes

$$\mu_{ij}(\beta, \gamma) = h(X_{ij}^T \beta + B(U_{ij})^T \gamma), \tag{2}$$

Denote $\theta = (\theta_1, \dots, \theta_p, \theta_{p+1}, \dots, \theta_{p+L})^T = (\beta^T, \gamma^T)^T$, $Y_i = (Y_{i1}, \dots, Y_{im_i})^T$, and write $X_i, U_i, B(U_i), \mu_i$ in a similar fashion. Following the QIFs approach, the extend score $g_N(\theta)$ is defined to be

$$g_N(\theta) = \frac{1}{n} \sum_{i=1}^n g_i(\theta) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \dot{\mu}_i^T A_i^{-1/2} M_1 A_i^{-1/2} (Y_i - \mu_i) \\ \vdots \\ \dot{\mu}_i^T A_i^{-1/2} M_s A_i^{-1/2} (Y_i - \mu_i) \end{pmatrix}, \tag{3}$$

where $\dot{\mu}_i = \frac{\partial \mu_i}{\partial \theta}$, $A_i = \text{diag}(v(\mu_{i1}), \dots, v(\mu_{im_i}))$ is the marginal covariance matrix of the i th subject and M_1, \dots, M_s are known matrices for approximating the in-

verse of the working correlation matrix $R(\rho)$ in GEEs. Then $\hat{\theta}$ is obtained by minimizing the following quadratic inference function

$$Q_n(\theta) = n g_N^T(\theta) \Omega_N(\theta)^{-1} g_N(\theta), \tag{4}$$

where $\Omega_N(\theta) = \frac{1}{n} \sum_{i=1}^n g_i(\theta) g_i(\theta)^T$.

Hence, $\hat{\beta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^T$ is the QIF estimator of β , and the estimator of $\alpha(u)$ can be obtained by $\hat{\alpha}(u) = B(u)^T \hat{\gamma}$, where $\hat{\gamma} = (\hat{\theta}_{p+1}, \dots, \hat{\theta}_{p+L})^T$ is the QIF estimator of γ . Details of the QIFs estimator for GPLMs with longitudinal data refers to [21].

2.2. GEL for GPLMs with Longitudinal Data

In most applications of GPLMs, the main interest is the statistical inference on the regression coefficient β_0 . Similar with [5], we regard the nonparametric function $\alpha(\cdot)$, i.e. the spline coefficient γ as nuisance, and conduct a suitable estimator of it to make sure the efficient statistical inference for β . In this article, we take the QIF estimate $\hat{\gamma}$ as the estimator of γ .

Noticing $g_i(\theta)$ in (3) carries the within-cluster correlation information, in order to construct the empirical likelihood ratio function for β , we introduce the auxiliary random vector

$$g_i(\beta; \hat{\gamma}) = \begin{pmatrix} \dot{\mu}_i(\beta, \hat{\gamma})^T A_i^{-1/2} M_i A_i^{-1/2} (Y_i - \mu_i(\beta, \hat{\gamma})) \\ \vdots \\ \dot{\mu}_i(\beta, \hat{\gamma})^T A_i^{-1/2} M_s A_i^{-1/2} (Y_i - \mu_i(\beta, \hat{\gamma})) \end{pmatrix}. \tag{5}$$

Note that $E(g_i(\beta; \hat{\gamma})) = 0$ when $\beta = \beta_0$, we define the generalized empirical log-likelihood ratio function as follows,

$$R(\beta; \hat{\gamma}) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_i(\beta; \hat{\gamma}) = 0 \right\}. \tag{6}$$

By the Lagrange multiplier method, we obtain that $R(\beta; \hat{\gamma})$ is maximized at

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T g_i(\beta; \hat{\gamma})} \tag{7}$$

where λ is a $ps \times 1$ vector satisfies

$$\frac{1}{n} \sum_{i=1}^n \frac{g_i(\beta; \hat{\gamma})}{1 + \lambda^T g_i(\beta; \hat{\gamma})} = 0. \tag{8}$$

Then $R(\beta; \hat{\gamma})$ can be represented as

$$R(\beta; \hat{\gamma}) = 2 \sum_{i=1}^n \log(1 + \lambda^T g_i(\beta; \hat{\gamma})). \tag{9}$$

By minimizing $R(\beta; \hat{\gamma})$ under the Equation constraints (8), we can obtain the maximum empirical likelihood estimator (MELE) of the parameter β .

3. Asymptotic Properties

For convenience and simplicity, let C denote a positive constant that may have different values at each appearance throughout this paper and $\|A\|$ denote the modulus of the largest singular value of matrix or vector A . Before the proof of our main theorems, we list some regularity conditions that used in this paper.

Assumption (A1): The spline regression parameter γ is identifiable, that is, γ_0 is the spline coefficient vector from the spline approximation to $\alpha_0(\cdot)$. In addition, there is a unique $\beta_0 \in S$ satisfying $E\{g_i(\beta, \hat{\gamma})\} = 0$, where S is the parameter space.

Assumption (A2): The weight matrix $\Omega_n = \frac{1}{n} \sum_{i=1}^n g_i(\theta) g_i^T(\theta)$ converges almost surely to a constant matrix Ω_0 , where Ω_0 is invertible.

Assumption (A3): The covariate matrices $X_i, i = 1, \dots, n$ satisfy that $\sup_i E\|X_i\|^4 < \infty$.

Assumption (A4): The error $\varepsilon_i = Y_i - \mu_i$ satisfies that $E(\varepsilon_i \varepsilon_i^T) = V_i, \sup_i \|V_i\| < \infty$, and there exists a positive constant δ such that $\sup_i E\|\varepsilon_i\|^{2+\delta} < \infty$.

Assumption (A5): All marginal variances $A_i \geq 0$ and $\sup_i \|A_i\| < \infty$.

Assumption (A6): $\{m_i\}$ is a bounded sequence of positive integers.

Assumption (A7): $\alpha(u)$ is r th continuous differentiable on $(0, 1)$, where $r \geq 2$.

Assumption (A8): The inner knots $\{c_i, i = 1, \dots, K\}$ satisfy

$$\max_{1 \leq i \leq K} |h_{i+1} - h_i| = o(K^{-1}) \text{ and } \frac{\max h_i}{\min h_i} \leq C_0,$$

where

$$h_i = c_i - c_{i-1}.$$

Assumption (A9): The link function $h(\cdot)$ is 2nd continuous differentiable and $E\{h^{2+\delta}\} < \infty$ for some $\delta > 2$.

[Remark] (A1) is for the identification. (A2) holds when based on the weak law of large numbers when n goes to infinity and the maximum cluster size is fixed, *i.e.*, when (A6) holds. (A3)-(A6) are the common regularity conditions in the longitudinal data analysis. (A7) is the usual assumption in spline approximation, it determines the convergence rate of spline estimate $\hat{\alpha} = B(U)^T \hat{\gamma}$. (A8) is the general condition for the knots in B-spline approximation. (A9) is the common condition in the study of GLMs.

We next study the asymptotic properties of the resulting GEL estimators. We first introduce some notations. Let β_0 denote the true values of $\beta(\cdot)$ and $\tilde{\beta}$ be the MELE of β . The following Theorem 1 shows that the $R(\beta; \hat{\gamma})$ is asymptotically distributed as a Chi-square with ps degrees of freedom.

Lemma 1. *Suppose that the regularity conditions of (A1)-(A9) hold and the numbers of knots $K = O(N^{1/(2r+1)})$, then*

$$\|\hat{\gamma} - \gamma_0\| = O_p(n^{-1/2}).$$

This is a direct result from Theorem 1 of [21].

Lemma 2. *Suppose that the regularity conditions of (A1)-(A9) hold and the numbers of knots $K = O(N^{1/(2r+1)})$, then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) \xrightarrow{L} N(0, \Omega_0). \tag{10}$$

$$\max_i \|g_i(\beta_0, \hat{\gamma})\| = o_p(n^{1/2}). \tag{11}$$

The proof can be found in the Appendix.

Theorem 1. *Assume that the conditions (A1)-(A9) hold and the numbers of knots $K = O(N^{1/(2r+1)})$, then*

$$R(\beta_0; \hat{\gamma}) \xrightarrow{L} \chi_{ps}^2,$$

where \xrightarrow{L} represents the convergence in distribution.

The proof can be found in the Appendix.

Let $\chi_{ps}^2(1-\alpha)$ be the $1-\alpha$ quantile of χ_{ps}^2 for any $0 < \alpha < 1$. From Theorem 1, an approximate $1-\alpha$ confidence region can be established by

$$C_\alpha(\beta) = \{R(\beta, \hat{\gamma}) \leq \chi_{ps}^2(1-\alpha)\}.$$

Denote

$$\Gamma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left(\frac{\partial g_i(\beta, \hat{\gamma})}{\partial \beta} \right),$$

$$\Omega_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left[g_i(\beta, \hat{\gamma}) g_i(\beta, \hat{\gamma})^T \right].$$

If the matrices Ω_0 and $\Gamma^T \Omega_0^{-1} \Gamma$ are invertible, we obtain the asymptotic normality of $\tilde{\beta}$.

Theorem 2. *Suppose that the conditions (A1)-(A9) hold and the numbers of knots $K = O(N^{1/(2r+1)})$, then*

$$\sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{L} N(0, V),$$

where $V = (\Gamma^T \Omega_0^{-1} \Gamma)^{-1}$.

The proof can be found in the Appendix.

The confidence region interval of each component of β is also worth concerning. Let e_r denote the unit vector with 1 at the r th entry, for $r = 1, \dots, p$. The estimate of the r th component of β is $\hat{\beta}_r = e_r^T \hat{\beta}$, for $r = 1, \dots, p$. Let

$$g_{ir}(\beta_r; \hat{\gamma}) = I_s \otimes e_r^T g_i(\hat{\beta}_1, \dots, \hat{\beta}_{r-1}, \beta_r, \hat{\beta}_{r+1}, \dots, \beta_p; \hat{\gamma})$$

where I_s is the $s \times s$ identity matrix, \otimes is the Kronecker product, and $g_i(\cdot)$ is defined in (5). Then, the partial generalized empirical log-likelihood ratio for β_r is defined as

$$R_r(\beta_r; \hat{\gamma}) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_{ir}(\beta; \hat{\gamma}) = 0 \right\}.$$

Theorem 3. *Assume that the conditions (A1)-(A9) and the numbers of knots $K = O(N^{1/(2r+1)})$, if β is the true parameter, then*

$$R_r(\beta_r; \hat{\gamma}) \xrightarrow{L} \chi_s^2.$$

The proof of Theorem 3 is similar to that of Theorem 1, we hence omit here.

Applying Theorem 3, the approximate $1 - \alpha$ confidence interval for β_r can be constructed by

$$C_\alpha(\beta_r) = \{R_r(\beta_r, \hat{\gamma}) \leq \chi_s^2(1 - \alpha)\}.$$

4. Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of the proposed methods. We compare the GEL with the NA-based method in terms of the coverage probability and the lengths of the obtained confidence region.

In our non-parametric estimation implementation, we use the sample quantiles of U_{ij} as knots. Moreover, we use cubic splines and take the number of internal knots to be the integer around $n^{1/5}$. This particular choice is consistent with the asymptotic theory in Section 3 and performs well in the simulations.

4.1. Study 1

Consider a binomial response:

$$\text{logit}\{Y_{ij} = 1 | X_{ij}, U_{ij}\} = X_{ij}\beta + 0.4 \cos(U_{2ij}\pi/2), i = 1, \dots, n; j = 1, \dots, m,$$

where $\beta = 0.4, m = 6, n = 50, 100$ and $150, X_{ij} \sim N(0, 1), U_{ij} \sim U[-1, 1]$. The clustered binary responses are generated as [23]. The correlation parameter ρ are taken to be 0.25, 0.5 and 0.75 which represent weak, medium and strong correlation respectively. We generated 500 data sets for each pair of (n, ρ) .

Table 1 list the EL-based and NA-based confidence intervals of β under CS structure. It shows that the GEL approach gives a slightly shorter confidence intervals than the NA method, while the former has a coverage probability more

Table 1. The average length and the corresponding coverage probabilities of the 95% confidence region of β for GEL and NA when the correlation structure is CS.

ρ	n	Average Length		Coverage Probability	
		GEL	NA	GEL	NA
0.25	50	0.3498	0.4821	0.932	0.834
	100	0.3191	0.3431	0.934	0.902
	150	0.2881	0.2795	0.958	0.928
0.5	50	0.3403	0.4850	0.919	0.824
	100	0.3159	0.3441	0.940	0.896
	150	0.2868	0.2792	0.956	0.922
0.75	50	0.3450	0.4829	0.904	0.811
	100	0.3166	0.3446	0.944	0.884
	150	0.2867	0.2792	0.944	0.908

closer to the nominal level. In addition, the coverage probability obtained by GEL approach tend to the nominal level and the average length decrease as n increases.

To study the influence of mis-specification to GEL approach, we derive the GEL confidence interval when the working correlation structure is specified to be CS and AR-1 respectively. **Table 2** list the results when the true structure is CS. **Table 3** list the results when the true structure is AR-1. It is known that the QIFs estimator is insensitive to mis-specification in correlation structure. **Table 2** and **Table 3** show that the QIFs-based GEL approach gives similar 95% confidence interval and coverage probability even the correlation structure is misspecified, which means the proposed GEL approach is robust.

Table 2. The average length and the corresponding coverage probabilities of the 95% confidence region of β for GEL when the true correlation structure is CS.

ρ	n	Average Length		Coverage Probability	
		CS	AR-1	CS	AR-1
0.25	50	0.3498	0.3544	0.932	0.920
	100	0.3191	0.3278	0.934	0.912
	150	0.2881	0.3025	0.958	0.966
0.5	50	0.3403	0.3370	0.919	0.919
	100	0.3159	0.3182	0.940	0.940
	150	0.2868	0.2867	0.956	0.956
0.75	50	0.3450	0.3427	0.904	0.916
	100	0.3166	0.3169	0.944	0.936
	150	0.2867	0.2795	0.944	0.938

Table 3. The average length and the corresponding coverage probabilities of the 95% confidence region of β for GEL when the true correlation structure is AR-1.

ρ	n	Average Length		Coverage Probability	
		CS	AR-1	CS	AR-1
0.25	50	0.3496	0.3501	0.940	0.904
	100	0.3177	0.3269	0.944	0.940
	150	0.2864	0.3008	0.958	0.944
0.5	50	0.3519	0.3599	0.942	0.922
	100	0.3153	0.3252	0.938	0.932
	150	0.2883	0.2915	0.942	0.946
0.75	50	0.3421	0.3201	0.948	0.920
	100	0.3092	0.3201	0.938	0.940
	150	0.2773	0.2852	0.944	0.942

4.2. Study 2

We consider a two-dimensional logistic model with $n = 150, m_i = m = 6$, and

$$\text{logit}\{Y_{ij} = 1 | X_{ij}, U_{ij}\} = X_{1ij}\beta_1 + X_{2ij}\beta_2 + 0.4\cos(U_{ij}\pi/2),$$

where $\beta = (\beta_1, \beta_2) = (0.5, 0.5)$, $X_{1ij} \sim N(0, 1)$, $X_{2ij} \sim N(0, 1)$ and U_{ij} are drawn independently from a uniform distribution on $[-1, 1]$. The clustered binary responses with exchangeable correlation structure are also generated as [23]. The correlation parameter ρ are taken to be 0.3 and 0.8.

Carried out 200 simulation runs, the EL-based and NA-based 95% confidence intervals for β when $\rho = 0.3$ are reported in **Figure 1**. It shows that GEL approach gives a smaller confidence region than the NA method. As to the coverage probability, the GEL approach is more closer to the nominal level than NA (0.925 vs 0.90). The result of $\rho = 0.8$ is similar, we omit here.

5. Example: Infectious Disease Data

To investigate the performance of the proposed method, we analysis an infectious disease data. In this study, a total of 275 preschool children were examined every three months for 18 months. The outcome is the presence of respiratory infection (1 = yes, 0 = no). The primary interest is in studying the relationship of the risk of respiratory infection to Vitamin A deficiency, which is indicated by xerophthalmia variable (1 = yes, 0 = no). The other covariates included: age,

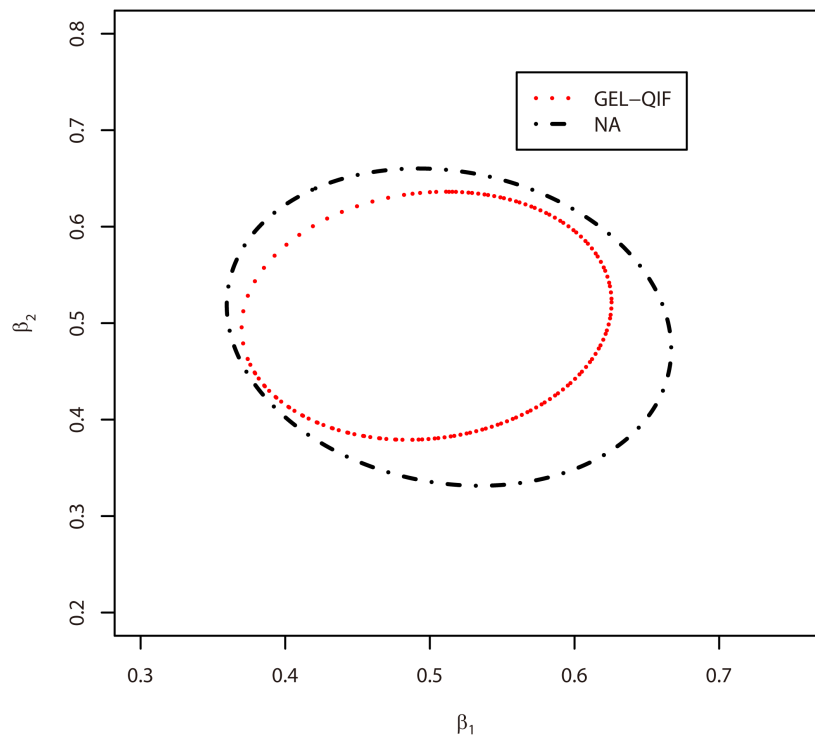


Figure 1. The 95% confidence region of β based on GEL and NA approach when $\rho = 0.3$ for Study 2.

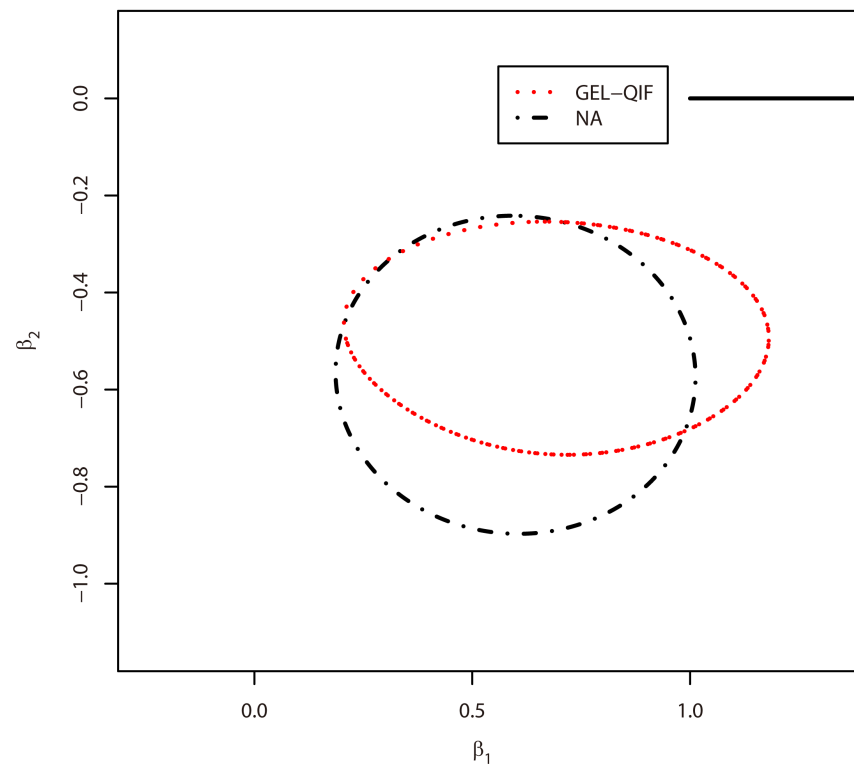


Figure 2. The 95% confidence region of β based on GEL and NA approach under exchangeable correlation structure for infectious disease data.

gender (1 = female, 0 = male), height, stunting status (1 = yes, 0 = no), and the seasonal Cosine and seasonal sine variables which indicate the season when those examinations took.

This data set has been well analyzed by many authors, such as [24] [25] [26] [27] [28]. We here consider a simple logistic model:

$$\text{logit}(\mu_{ij}) = \beta_1 X_{1ij} + \beta_2 X_{2ij} + \alpha(U_{ij}),$$

where μ is the mean of the risk of respiratory infection, β_1 and β_2 describe the effects of Vitamin A deficiency and the sex aspect. We use two methods: the NA method and QIFs-based GEL under the CS correlation. The confidence regions are reported in **Figure 2**. It shows that GEL gives smaller confidence regions than NA does.

Acknowledgements

We thank the Editor and the referees for their comments. The research is funded by the National Natural Science Foundation of China (11571025) and the Beijing Natural Science Foundation (1182008). This support is greatly appreciated.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Owen, A. (1988) Empirical Likelihood Ratio Confidence Intervals for a Single Function. *Biometrika*, **75**, 237-249. <https://doi.org/10.1093/biomet/75.2.237>
- [2] Owen, A. (1990) Empirical Likelihood and Confidence Regions. *Annals of Statistics*, **18**, 90-120. <https://doi.org/10.1214/aos/1176347494>
- [3] Qin, J. and Lawless, J.F. (1994) Empirical Likelihood and General Estimating Equations. *Annals of Statistics*, **22**, 300-325. <https://doi.org/10.1214/aos/1176325370>
- [4] Owen, A. (2001) Empirical Likelihood. Chapman & Hall, London. <https://doi.org/10.1201/9781420036152>
- [5] Shi, J. and Lau, T.S. (2000) Empirical Likelihood for Partially Linear Models. *Journal of Multivariate Analysis*, **72**, 132-148. <https://doi.org/10.1006/jmva.1999.1866>
- [6] Wang, Q.H. and Jing, B. (2003) Empirical Likelihood for Partially Linear Models. *Annals of the Institute of Statistical Mathematics*, **55**, 585-595. <https://doi.org/10.1007/BF02517809>
- [7] Xue, L.G. and Zhu, L.X. (2007) Empirical Likelihood Semiparametric Regression Analysis for Longitudinal. *Biometrika*, **94**, 921-937. <https://doi.org/10.1093/biomet/asm066>
- [8] Xue, L.G. and Zhu, L.X. (2008) Empirical Likelihood-Based Inference in a Partially Linear Model for Longitudinal Data. *Science in China, Ser A*, **51**, 115-130. <https://doi.org/10.1007/s11425-008-0020-4>
- [9] Bai, Y., Fung, W.K. and Zhu, Z.Y. (2010) Weighted Empirical Likelihood for Generalized Linear Models with Longitudinal Data. *Journal of Statistical Planning and Inference*, **140**, 3446-3456. <https://doi.org/10.1016/j.jspi.2010.05.007>
- [10] Qu, G.Y., Bai, Y. and Zhu, Z.Y. (2012) Robust Empirical Likelihood Inference for Generalized Partial Linear Models with Longitudinal Data. *Journal of Multivariate Analysis*, **105**, 21-44. <https://doi.org/10.1016/j.jmva.2011.08.003>
- [11] Tian, R.Q., Xue, L.G. and Liu, C.L. (2014) Generalized Empirical Likelihood Inference in Generalized Linear Models for Longitudinal Data. *Communications in Statistics—Theory and Methods*, **43**, 3893-3904. <https://doi.org/10.1080/03610926.2012.719987>
- [12] Tian, R.Q. and Xue, L.G. (2017) Generalized Empirical Likelihood Inference in Partial Linear Models for Longitudinal Data. *Statistics*, **51**, 988-1005. <https://doi.org/10.1080/02331888.2017.1355370>
- [13] Liang, K.L. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Estimating Equations. *Biometrika*, **73**, 13-22. <https://doi.org/10.1093/biomet/73.1.13>
- [14] Qu, A., Lindsay, B.G. and Li, B. (2000) Improving Generalized Estimating Equations Using Quadratic Inference Functions. *Biometrika*, **87**, 823-836. <https://doi.org/10.1093/biomet/87.4.823>
- [15] Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, **50**, 1029-1054. <https://doi.org/10.2307/1912775>
- [16] Qu, A. and Song, X.K. (2004) Assessing Robustness of Generalized Estimating Equations and Quadratic Inference Functions. *Biometrika*, **91**, 447-459. <https://doi.org/10.1093/biomet/91.2.447>
- [17] Qu, A. and Li, R. (2006) Quadratic Inference Functions for Varying Coefficient Models with Longitudinal Data. *Biometrics*, **62**, 379-391. <https://doi.org/10.1111/j.1541-0420.2005.00490.x>

- [18] Zhou, J.L. and Qu, A. (2012) Informative Selection for Correlation Structure for Longitudinal Data. *Journal of the American Statistical Association*, **107**, 701-710. <https://doi.org/10.1080/01621459.2012.682534>
- [19] Wang, L., Xue, L., Qu, A. and Liang, H. (2014) Estimation and Model Selection in Generalized Additive Partial Linear Models for Correlated Data with Diverging Number of Covariates. *Annals of Statistics*, **42**, 592-694. <https://doi.org/10.1214/13-AOS1194>
- [20] Tian, R.Q., Xue, L.G. and Liu, C.L. (2014) Penalized Quadratic Functions for Semiparametric Varying Coefficient Partially Linear Models with Longitudinal Data. *Journal of Multivariate Analysis*, **132**, 94-110. <https://doi.org/10.1016/j.jmva.2014.07.015>
- [21] Zhang, J.H. and Xue, L.G. (2017) Quadratic Inference Functions for Generalized Partially Linear Models with Longitudinal Data. *Chinese Journal of Applied Probability and Statistics*, **33**, 409-416. <http://aps.ecnu.edu.cn/EN/10.3969/j.issn.1001-4268.2017.04.007>
- [22] He, X., Zhu, Z.Y. and Fung, W.K. (2002) Estimation in a Semiparametric Model for Longitudinal Data with Unspecified Dependence Structure. *Biometrika*, **89**, 579-590. <https://doi.org/10.1093/biomet/89.3.579>
- [23] Oman, S.D. (2009) Easily Simulated Multivariate Binary Distributions with Given Positive and Negative Correlations. *Computational Statistics & Data Analysis*, **53**, 999-1005. <https://doi.org/10.1016/j.csda.2008.11.017>
- [24] Lin, X.H. and Carroll, R.J. (2001) Semiparametric Regression for Clustered Data. *Biometrika*, **88**, 1179-1185. <https://doi.org/10.1093/biomet/88.4.1179>
- [25] Lin, X.H. and Carroll, R.J. (2001) Semiparametric Regression for Clustered Data Using Generalized Estimating Equations. *Journal of the American Statistical Association*, **96**, 1045-1056. <https://doi.org/10.1198/016214501753208708>
- [26] He, X.M., Fung, W.K. and Zhu, Z.Y. (2005) Robust Estimation in a Generalized Partially Linear Model for Cluster Data. *Journal of the American Statistical Association*, **100**, 1176-1184. <http://doi.org/10.1198/016214505000000277>
- [27] Zeger, S.L. and Karim, M.R. (1991) Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. *Journal of the American Statistical Association*, **86**, 79-86. <https://doi.org/10.1080/01621459.1991.10475006>
- [28] Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994) *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
- [29] Schumaker, G. (1981) *Spline Function*. Wiley, New York.

Appendix

Proof of Lemma 2

Proof. Consider the k th component of $g_N(\beta_0, \hat{\gamma})$:

$$g_{N,k}(\beta_0) = \frac{1}{n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} (Y_i - \mu_i(\beta_0, \hat{\gamma})). \quad (12)$$

Note that

$$Y_i - \mu_i(\beta_0, \hat{\gamma}) = h(X_i \beta_0 + \alpha_0(U_i)) - h(X_i \beta_0 + B(U_i) \hat{\gamma}) + \varepsilon_i, \quad (13)$$

Apply Taylor expansion to the first two terms in (13) at $\eta_i = X_i \beta_0 + B(U_i) \gamma_0$, we obtain

$$Y_i - \mu_i(\beta_0, \hat{\gamma}) = \varepsilon_i + H'(\eta_i) R_1(U_i) + \frac{1}{2} H''(\omega_i) R_1^2(U_i) - H'(\eta_i) R_2(U_i) - \frac{1}{2} H''(\pi_i) R_2^2(U_i), \quad (14)$$

where ω_i is between η_i and $X_i \beta_0 + \alpha_0(U_i)$, π_i is between η_i and $X_i \beta_0 + B(U_i) \hat{\gamma}$, and

$$\begin{aligned} R_1(U_{ij}) &= \alpha_0(U_{ij}) - B(U_{ij})^T \gamma_0, \quad R_1(U_i) = (R_1(U_{i1}), \dots, R_1(U_{im}))^T, \\ R_2(U_{ij}) &= B(U_{ij})^T (\hat{\gamma} - \gamma_0), \quad R_2(U_i) = (R_2(U_{i1}), \dots, R_2(U_{im}))^T, \\ R_1^2(U_i) &= \left([R_1(U_{i1})]^2, \dots, [R_1(U_{im})]^2 \right)^T, \quad R_2^2(U_i) = \left([R_2(U_{i1})]^2, \dots, [R_2(U_{im})]^2 \right)^T, \\ H'(\eta_i) &= \text{diag}(h'(\eta_{i1}), \dots, h'(\eta_{im})), \quad H''(\eta_i) = \text{diag}(h''(\eta_{i1}), \dots, h''(\eta_{im})). \end{aligned}$$

Substitute (14) into (12), we obtain

$$\begin{aligned} g_{N,k}(\beta_0) &= \frac{1}{n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} \varepsilon_i \\ &\quad + \frac{1}{n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} H'(\eta_i) R_1(U_i) \\ &\quad + \frac{1}{2n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} H''(\omega_i) R_1^2(U_i) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} H'(\eta_i) R_2(U_i) \\ &\quad - \frac{1}{2n} \sum_{i=1}^n \dot{\mu}_i^T A_i^{-1/2} M_k A_i^{-1/2} H''(\pi_i) R_2^2(U_i) \\ &\equiv I_1 + I_2 + I_3 + I_4 + I_5. \end{aligned}$$

From conditions (A7), (A8) and theorem 12.7 in [29], we have $\|R_1(U_{ij})\| = O(K^{-r})$ and $\|B(\cdot)\| = O(1)$. Then, invoking conditions (A3)-(A5), by a simple calculation, we have $I_2 = O_p(n^{-1} n^{1/2} K^{-r}) = o_p(n^{-1/2})$ and $I_3 = o_p(n^{-1/2})$.

Invoking conditions (A4)-(A9), by lemma 1, we have $I_4 = O_p(n^{-1} n^{1/2} n^{-1/2}) = o_p(n^{-1/2})$ and $I_5 = o_p(n^{-1/2})$.

Denote

$$\begin{aligned} \zeta_{ik} &= \mu_i^T A_i^{-1/2} M_k A_i^{-1/2} \varepsilon_i, \\ \xi_i &= (\zeta_{i1}^T, \dots, \zeta_{is}^T)^T, \end{aligned}$$

we have

$$g_N(\beta_0, \hat{\gamma}) = \frac{1}{n} \sum_{i=1}^n \xi_i + o_p(n^{-1/2}).$$

Follow [11], we obtain

$$\sqrt{n} g_N(\beta_0, \hat{\gamma}) \xrightarrow{\mathcal{L}} N(0, \Omega_0).$$

i.e.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) \xrightarrow{\mathcal{L}} N(0, \Omega_0).$$

Similarly, (11) can be proved. Thus we complete the proof of the Lemma 2.

Follow the argument of [4], we can prove

$$\|\lambda\| = O_p(n^{1/2}). \tag{15}$$

Proof of Theorem 1

Proof. Applying Taylor expansion to $R(\beta_0, \hat{\gamma})$, we obtain

$$R(\beta_0, \hat{\gamma}) = 2 \sum_{i=1}^n \left\{ \lambda^T g_i(\beta_0, \hat{\gamma}) - \left[\lambda^T g_i(\beta_0, \hat{\gamma}) \right]^2 / 2 \right\} + o_p(1).$$

Recall (8), it follows that

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{g_i(\beta; \hat{\gamma})}{1 + \lambda^T g_i(\beta; \hat{\gamma})} \\ &= \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) - \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) g_i^T(\beta_0, \hat{\gamma}) \lambda \\ &\quad + \sum_{i=1}^n \frac{g_i(\beta_0, \hat{\gamma}) \left[\lambda^T g_i(\beta_0, \hat{\gamma}) \right]^2}{1 + \lambda^T g_i(\beta_0, \hat{\gamma})}. \end{aligned}$$

This together with Lemma 1 and (15) proves that

$$\sum_{i=1}^n \left[\lambda^T g_i(\beta_0, \hat{\gamma}) \right]^2 = \sum_{i=1}^n \lambda^T g_i(\beta_0, \hat{\gamma}) + o_p(1),$$

and

$$\lambda = \left\{ \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) g_i^T(\beta_0, \hat{\gamma}) \right\}^{-1} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) + o_p(n^{-1/2}).$$

Therefore, we have

$$\begin{aligned} R(\beta_0, \hat{\gamma}) &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) \right\}^T \left\{ \frac{1}{n} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) g_i^T(\beta_0, \hat{\gamma}) \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta_0, \hat{\gamma}) \right\} + o_p(1). \end{aligned}$$

Together with Lemma 2, we complete the proof of Theorem 1.

Proof of Theorem 2

Proof. We first define bivariate functions $Q_1(\beta, \lambda, \hat{\gamma})$ and $Q_2(\beta, \lambda, \hat{\gamma})$ respectively as

$$Q_{1n}(\beta, \lambda, \hat{\gamma}) = \frac{1}{n} \sum_{i=1}^n \frac{g_i(\beta, \hat{\gamma})}{1 + \lambda^T g_i(\beta, \hat{\gamma})},$$

$$Q_{2n}(\beta, \lambda, \hat{\gamma}) = \frac{1}{n} \sum_{i=1}^n \frac{g_i(\beta, \hat{\gamma})}{1 + \lambda^T g_i(\beta, \hat{\gamma})} \left(\frac{\partial(g_i(\beta))}{\partial \beta} \right)^T \lambda.$$

Under the condition (A1), if $\tilde{\beta}$ is the MELE of β and $\tilde{\lambda}$ is the root of (8), following Lemma 1 of [3], we have

$$Q_{1n}(\tilde{\beta}, \tilde{\lambda}, \hat{\gamma}) = 0, \quad Q_{2n}(\tilde{\beta}, \tilde{\lambda}, \hat{\gamma}) = 0.$$

Expanding $Q_{1n}(\tilde{\beta}, \tilde{\lambda}, \hat{\gamma})$ and $Q_{2n}(\tilde{\beta}, \tilde{\lambda}, \hat{\gamma})$ at $(\beta_0, \tilde{\lambda}; \hat{\gamma})$, together with conditions (A6)-(A9) and Lemma 2, we have

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{P} (\Gamma^T \Omega_0^{-1} \Gamma)^{-1} \Gamma^T \Omega_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i(\beta, \hat{\gamma}) \xrightarrow{L} N(0, V).$$