Journal of Multivariate Analysis 114 (2013) 63–73

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Empirical likelihood for generalized linear models with longitudinal data

Daoji Li, Jianxin Pan *

School of Mathematics, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

ARTICLE INFO

Article history:
Received 20 June 2011
Available online 24 July 2012

AMS subject classifications:
62H12
62H15
62H20
62J12

Keywords:
Empirical likelihood
Generalized estimating equations
Longitudinal data
Quadratic inference functions
Quasi-likelihood
Hypothesis testing

ABSTRACT

In this paper, empirical likelihood-based inference for longitudinal data within the framework of generalized linear model is investigated. The proposed procedure takes into account the within-subject correlation without involving direct estimation of nuisance parameters in the correlation matrix and retains optimal even if the working correlation structure is misspecified. The proposed approach yields more efficient estimators than conventional generalized estimating equations and achieves the same asymptotic variance as quadratic inference function based methods. Furthermore, hypothesis testing procedures are developed to test whether or not the model assumption is met and whether or not regression coefficients are significant. The finite sample performance of the proposed methods is evaluated through simulation studies. Application to the Ohio Children Wheeze Status data is also discussed.

1. Introduction

Longitudinal data, or more generally cluster data, often arise in the biometrical, epidemiological, social and economical fields. Repeated measurements are made on subjects over time and responses within a subject are very likely to be correlated with an unknown correlation structure, although responses between subjects may be independent. When modeling longitudinal data, the within-subject correlation must be taken into account. Otherwise, statistical inference may not be correct [7].

The method of generalized estimating equations (GEEs), developed by Liang and Zeger [17] from generalized linear models [21,20] and quasi-likelihood [35,19], is widely used to deal with both continuous and discrete longitudinal data. The key idea behind the GEE approach is to extend generalized linear models and quasi-likelihood methods by using a "working" correlation matrix parameterized in terms of additional nuisance parameters α, with additional estimating equations for α if necessary. The estimators of parameters in the mean based on the GEE approach are still consistent even if the correlation matrix is misspecified [17].

However, when the working correlation is misspecified, this can lead to a great loss of efficiency of parameter estimators. Furthermore, one of the underlying assumptions on the GEE method is that the nuisance parameter α in the correlation matrix has to be properly estimated. Unfortunately, the estimator of α suggested by Liang and Zeger [17] is not available even in some simple cases of misspecification [5].

The generalized method of moments (GMM), proposed by Hansen [10], is a popular approach for estimation of the vector of regression parameters from a set of score functions when the dimension of the score function exceeds that of the regression parameter. Under some regularity conditions, the estimators obtained via GMM are consistent and asymptotically normal. To overcome the difficulties associated with the GEE method, Qu et al. [30] introduced a method based on quadratic

* Corresponding author.
E-mail addresses: d.li@maths.man.ac.uk (D. Li), jpan@maths.man.ac.uk, jianxin.pan@manchester.ac.uk (J. Pan).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.07.014

inference functions (QIF) which does not involve direct estimation of the nuisance parameter α but lead to more efficient estimators even if the working correlation matrix is misspecified. The key idea of QIF is to represent the inverse of the working correlation matrix by the linear combination of basis matrices and combine these estimating equations using the principle of GMM. More recently, the QIF approach has been applied to other models, including varying coefficient models [29], partial linear models [3], and single-index models [1].

An important problem in statistical inference is how to construct confidence regions for parameters of interest. Two conventional methods are the normal approximation method and the bootstrap method. A convenient choice is the use of asymptotic normal distribution of parameter estimators obtained by one of the aforementioned methods to construct the confidence regions. However, with this method, a plug-in estimator of the limiting variance of regression parameter estimators is needed. Accordingly, the confidence region that is derived from an asymptotic normal distribution is predetermined to be symmetric. To avoid plug-in estimation of the limiting variance, the bootstrap method may be used instead.

Taking these issues into account, we consider a different but somehow related approach via empirical likelihood (EL) to construct the confidence regions. The EL approach, originally proposed by Owen [23], has many advantages over the normal approximation-based method and the bootstrap method [6,9,23–26,28]. First, the EL approach does not involve a plug-in estimation for the limiting variance. Second, the shape and the orientation of EL-based confidence regions are determined automatically and entirely by the data. Third, the EL approach yields better coverage probability for small sample [26]. Fourth, as DiCiccio et al. [6] showed, the EL approach is Bartlett correctable, and thus has an advantage over the bootstrap method.

The EL approach has been widely applied to statistical models to construct a confidence region for the mean of a random vector in a similar manner to that by Owen [23,24]. Owen [25] considered the EL-based analysis for linear regression models and Kolaczyk [13] extended it to generalized linear models. Further extensions to partial linear models and generalized partially linear models were also considered [16,31,33]. However, all those aforementioned methods assume that data are statistically independent. The use of the EL-approach to longitudinal data analysis has received little attention because of the challenge imposed by incorporating the within-subject correlation. Existing work using the EL-approach for longitudinal data includes longitudinal partially linear regression models [40], case-control studies [42], varying coefficient models [36,37], semiparametric varying coefficient partially linear models [43,44,38] and partially linear single-index models [18]. All those studies, however, use a working independence structure and thus ignore the within-subject correlation of longitudinal data. Recently, Wang et al. [34] extended the block empirical likelihood method of You et al. [40] by accounting for the within-subject correlation and proposed two generalized EL-based methods, the subject-wise EL method and the element-wise EL method, to solve the problem. Wang et al. [34] showed that the subject-wise EL method with correctly specified correlation structure is the most efficient. Bai et al. [2] proposed a weighted EL-approach for longitudinal data by introducing a weight matrix to account for the within-subject correlation. However, it is still not clear yet how to correctly specify the correlation structure in [34] and how to choose the optimal weight matrix in [2].

In this paper, we aim to develop an unified approach that enables us to deal with both continuous and discrete longitudinal data without losing efficiency. We propose an estimation procedure for longitudinal data using a linear approximation to the inverse of the correlation matrix and empirical likelihood. The proposed method allows us to directly incorporate correlation into model building, but does not require estimation of the nuisance parameters associated with the correlation.

We also consider confidence regions for parameters of interest. The EL-based confidence regions for the unknown regression parameters can be constructed. Our proposed approach does not require estimating the covariance matrices of the parameter estimators. Moreover, the proposed confidence region is adapted to data and not necessarily symmetric. Thus, it reflects the nature of the underlying data and hence gives a more representative manner to make inferences about the parameters of interest.

The rest of the paper is organized as follows. In Section 2, we first briefly review the GEE and QIF approaches. We then define an EL ratio function based on the "extended score" and propose an EL-estimation approach for longitudinal data, which can be obtained by using a linear approximation in the line of the QIF approach. Section 3 discusses the theoretical results of the proposed estimation method. Hypothesis testing procedures for the model assumption and regression coefficient specification are also developed. Section 4 reports the finite sample performance of the proposed method by Monte Carlo simulation studies. Section 5 gives an application of the proposed estimation method to the Ohio Children Wheeze Status data. Further discussions are given in Section 5.

2. Estimation procedure

In this section, we describe how to obtain efficient estimators of parameters using EL. We start with a brief review on existing methods for longitudinal data under the framework of generalized linear models.

2.1. Generalized estimating equations

Let y_{ij} be an outcome variable and x_{ij} be a $p \times 1$ vector of covariates, observed at times t_{ij} ($j = 1, \ldots, m_i$) for subjects $i = 1, \ldots, n$. Assume that the first and second moments of y_{ij} are modeled by

$$g(\mu_{ij}) = x_{ij}^T\beta \tag{1}$$

and

$$Var(y_{ij}) = \phi v(\mu_{ij}) \tag{2}$$

where β is a $p \times 1$ parameter vector, $g(\cdot)$ is a known link function, $\mu_{ij} = E(y_{ij})$, ϕ is a dispersion parameter, and $v(\cdot)$ is a known variance function.

For the longitudinal data, the quasi-likelihood equation [35,19] of β is defined by

$$U(\beta) = \sum_{i=1}^{n}(\dot{\mu}_i)^T \Sigma_i^{-1}(y_i - \mu_i) = 0, \tag{3}$$

where $\dot{\mu}_i = \partial\mu_i/\partial\beta$, $\mu_i = (\mu_{i1}, \ldots, \mu_{im_i})^T$, $\Sigma_i = Cov(y_i)$ and $y_i = (y_{i1}, \ldots, y_{im_i})^T$. If Σ_i is known, it is well known that $U(\beta)$ is the optimal estimating function.

In practice, Σ_i is often unknown, so that the estimator of β is obtained by solving (3) after replacing Σ_i in (3) with empirical estimator $\hat{\Sigma}_i$ for the Σ_i. However, if the size of matrix Σ_i is large, the inverse of an estimator of Σ_i, e.g., based on sample covariance, can be very unreliable or incalculable, especially when Σ_i has many zero entries, as is common in spatial data [14], or if Σ_i is close to a singular matrix.

To avoid this difficulty, Liang and Zeger [17] proposed to keep the same assumption about the marginal variance structure as in quasi-likelihood equations but introduce a common working correlation matrix R which involves a small number of nuisance parameters α to simplify the correlation analysis. Specifically, it is assumed that the matrix Σ_i can be expressed in terms of a working correlation matrix $R(\alpha)$ as $\Sigma_i = A_i^{1/2}R(\alpha)A_i^{1/2}$, where $A_i = diag\{Var(y_{i1}), \ldots, Var(y_{im_i})\}$ and α is some unknown nuisance parameter. Thus, the corresponding GEE are

$$\sum_{i=1}^{n}(\dot{\mu}_i)^T A_i^{-1/2}R^{-1}(\alpha)A_i^{-1/2}(y_i - \mu_i) = 0. \tag{4}$$

If the working correlation R is misspecified, the resulting estimator of the parameters β based on GEE (4) is still consistent, but it may not be efficient within the same class of estimating equations [32].

As pointed out by Crowder [5], one potential limitation of the GEE method is that the moment estimator of the nuisance parameter α in the working correlation matrix may not exist even in some simple cases.

2.2. Quadratic inference function

Unlike the GEE, Qu et al. [30] model R^{-1} by the class of matrices

$$R^{-1} \approx \sum_{i=1}^{k} a_i M_i, \tag{5}$$

where $M_1 = I$ is the identity matrix and $M_2, \ldots, M_k$ are known symmetric basis matrices with 0 or 1 as their components and $a_1, \ldots, a_k$ are unknown constants determining the within-subject correlation parameters α, which can be viewed as nuisance parameters associated with the correlation. It is worth pointing out that the basis matrices $M_1, \ldots, M_k$ do not depend on the nuisance parameters α. This expression (5) holds exactly for some common working correlation structures. For example, if $R(\alpha)$ is a matrix with compound symmetry (CS) structure, then $R^{-1}(\alpha) = a_1M_1+a_2M_2$ with M_2 being a matrix with 0 on the diagonal and 1 off the diagonal. And if $R(\alpha)$ is a matrix with AR(1) structure, then $R^{-1}(\alpha) = a_1^*M_1+a_2^*M_2+a_3^*M_3$ where M_2 takes 1 on the two main off diagonals and zero elsewhere and M_3 takes 1 on the two corner components of the diagonal and zero elsewhere. In general, M_3 is a minor boundary correction and can be omitted. For more details, see [30].

Substituting (5) into (4) leads to

$$\sum_{i=1}^{n}(\dot{\mu}_i)^T A_i^{-1/2}(a_1M_1 + \cdots + a_kM_k)A_i^{-1/2}(y_i - \mu_i) = 0. \tag{6}$$

One approach would be to choose the coefficients $a = (a_1, \ldots, a_k)$ so as to optimize some function of the information matrix associated with (6). Note that the estimating equations (6) are a linear combination of elements of the following "extended score" vector

$$\bar{S}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} S_i(\beta),$$

where

$$S_i(\beta) = \begin{pmatrix} (\dot{\mu}_i)^T A_i^{-1/2}M_1A_i^{-1/2}(y_i - \mu_i) \\ \vdots \\ (\dot{\mu}_i)^T A_i^{-1/2}M_kA_i^{-1/2}(y_i - \mu_i) \end{pmatrix}. \tag{7}$$

In general, there is no solution for the equations $\bar{S}_n(\beta) = 0$ unless $k = 1$ as $\bar{S}_n(\beta)$ is a vector of length kp ($> p$), which means that there are more estimating equations than those actually needed for estimation of the unknown parameters.

Qu et al. [30] proposed to use the GMM [10] to construct an estimator of β by minimizing the quadratic inference function

$$Q_n(\beta) = n\bar{S}_n^T(\beta)\, C_n^{-1}(\beta)\, \bar{S}_n(\beta), \tag{8}$$

where $C_n(\beta) = n^{-1} \sum_{i=1}^n S_i(\beta) S_i^T(\beta)$ is the sample covariance matrix and is assumed to be invertible.

According to Hansen [10], the QIF estimator, i.e., $\hat{\beta} = \arg\min_\beta Q_n(\beta)$, has the usual large sample properties under some regularity conditions. In other words,

$$\sqrt{n}(\hat{\beta} - \beta) \longrightarrow N(0, V_{QIF}) \tag{9}$$

in distribution as $n \to \infty$, where the asymptotic covariance matrix $V_{QIF} = \left(\Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\right)^{-1}$ with

$$\Sigma_{12} = \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\beta)}{\partial \beta}\right] \quad \text{and} \quad \Sigma_{11} = \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^n S_i(\beta) S_i^T(\beta)\right]. \tag{10}$$

The asymptotic covariance matrix V_{QIF} can be estimated consistently by $\hat{V}_{QIF}$ that is given by

$$\left\{\left[\frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}}\right]^T \left[\frac{1}{n} \sum_{i=1}^n S_i(\hat{\beta}) S_i^T(\hat{\beta})\right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}}\right]\right\}^{-1}.$$

Based on these results, the approximated confidence regions are provided by

$$\left\{\beta\ : n(\hat{\beta} - \beta)^T \hat{V}_{QIF}(\hat{\beta} - \beta) \leqslant \chi^2_{p,\alpha}\right\}$$

where $\chi^2_{p,\alpha}$ is the $1 - \alpha$ quantile of the standard χ^2_p distribution.

If the dimension of the extended score is the same as the dimension of the parameters, that is, $k = 1$, then $n^{-1} \sum_{i=1}^n \partial S_i(\beta)/\partial\beta$ and $n^{-1} \sum_{i=1}^n S_i(\beta) S_i^T(\beta)$ are $p \times p$ matrices and minimizing the QIF is equivalent to solving $\bar{S}_n(\beta) = 0$ directly.

This approach also provides an optimal linear combination of the given estimating functions in the sense that the asymptotic variance of the estimator attains the minimum in the sense of Loewner ordering (e.g. [27, p. 12]). This property ensures that the QIF approach improves the efficiency of the GEE method when the working correlation structure is misspecified but remains as efficient as the GEE method when the working correlation structure is specified correctly [30].

2.3. Empirical likelihood

By representing the inverse of the working correlation matrix by a linear combination of the basis matrices and using only those basis matrices to formulate the objective function, the QIF estimator is obtained with no need to estimate the nuisance correlation parameters $a_1, \ldots, a_k$. Hence, the QIF method does not depend on whether or not appropriate estimators of the correlation parameters are available. As there are more estimating equations than those actually needed for estimation of the unknown parameters, the GMM is used through minimizing the criterion of the QIF.

GMM and EL are two popular but different methods for combining information about parameters when there are more estimating equations than those actually needed for estimation of the unknown parameters. The main difference lies in the way of combination of the information [28,11]. Imbens [12] and Newey and Smith [22] further showed that the EL-approach has several advantages over the GMM-method. For example, the GMM estimation requires a correct choice of the weight matrix. In contrast, the EL-approach does not have such a constraint. Moreover, the EL estimation can produce a smaller bias of second order than the GMM method. See [12,22] for more details. Below we carry out a likelihood-type inference using EL [23], encouraged by its attractive performance for estimating equations as demonstrated by Qin and Lawless [28] and others.

Let p_i be the non-negative probability weight associated with the estimating function $S_i(\beta)$ defined in (7). Note that $\{S_i(\beta) : 1 \le i \le n\}$ are mutually independent and satisfy $E[S_i(\beta_0)] = 0$ where β_0 is the true value of β. The EL ratio function for β is defined by

$$L_n(\beta) = \max\left\{\prod_{i=1}^n np_i \bigg| p_i \ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i S_i(\beta) = 0\right\}, \tag{11}$$

where the maximization is taken with respect to the probabilities $p_1, \ldots, p_n$. Note for a given β an unique value for $L_n(\beta)$ exists, provided that 0 is inside the convex hull of the point $S_1(\beta), \ldots, S_n(\beta)$ [23,24]. By the standard procedure of the EL approach [28,23], we know that the probability p_i must have the form

$$p_i = \frac{1}{n}\left[\frac{1}{1 + t^T(\beta) S_i(\beta)}\right],$$

where $t(\beta)$ is the Lagrange multiplier that satisfies

$$Q_{n1}(\beta,t)=\frac{1}{n}\sum_{i=1}^{n}\frac{S_i(\beta)}{1+t^T(\beta)S_i(\beta)}=0. \tag{12}$$

Clearly, we have

$$\log L_n(\beta)=-\sum_{i=1}^{n}\log\left[1+t^T(\beta)S_i(\beta)\right]. \tag{13}$$

The empirical log-likelihood ratio function of β can be defined by

$$\ell_n(\beta)=-\log L_n(\beta)=\sum_{i=1}^{n}\log\left[1+t^T(\beta)S_i(\beta)\right]. \tag{14}$$

Then the maximum EL estimator (MELE), $\tilde{\beta}$, is defined as the maximizer of $L_n(\beta)$ or the minimizer of $\ell_n(\beta)$.

When the estimating function $S_i(\beta)$ is differentiable with respect to β, the MELE can be found via solving the following system of equations,

$$Q_{n1}(\beta,t)=0 \quad\text{and}\quad Q_{n2}(\beta,t)=0, \tag{15}$$

where $Q_{n1}(\beta,t)$ is defined in (12) and

$$Q_{n2}(\beta,t)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+t^T(\beta)S_i(\beta)}\left\{\frac{\partial S_i(\beta)}{\partial\beta}\right\}^T t(\beta). \tag{16}$$

In general, $t(\beta)$ is a vector with dimension kp where k is the number of the basis matrices involved in the working correlation structure. As mentioned earlier, for some common working correlation structures k takes small integers, for example, $k=1$ for working independent correlation structure and $k=2$ for compound symmetry and AR(1) correlation structures. According to [28], when $k=1$, the EL estimator $\tilde{\beta}$ is also equal to the solution of the estimating equations

$$\sum_{i=1}^{n}S_i(\beta)=0$$

We show in the next section that if β_0 is the true parameter vector, $2\ell_n(\beta_0)$ is then asymptotically chi-square distributed.

3. Main results

In this section, we give asymptotic properties of the EL estimator $\tilde{\beta}$ and the empirical log-likelihood ratio. To state the results, we first introduce some notation. In the following, we use $\|\cdot\|$ to denote the Euclidean norm for a vector.

C1. The matrix $n^{-1}\sum_{i=1}^{n}S_iS_i^T$ converges almost surely to a constant matrix Σ_{11}, where Σ_{11} is positive definite.
C2. The covariates x_{ij} and the vectors

$$\dot{g}^{-1}(x_i\beta)=(\dot{g}^{-1}(x_{i1}'\beta),\ldots,\dot{g}^{-1}(x_{im_i}^T\beta))^T$$

are all bounded, meaning that all the elements of the vectors are bounded.
C3. $E\|y_i-\mu_i\|^3<\infty$.
C4. The matrix $n^{-1}\sum_{i=1}^{n}\partial S_i(\beta)/\partial\beta$ converges in probability to a constant matrix Σ_{12}.
C5. The parameter space S is a compact subset of R^p, and β_0 is an interior point of S.
C6. $\partial^2 S_i(\beta)/\partial\beta\partial\beta^T$ is continuous with respect to β in a neighborhood of the true value β_0.

Conditions listed in C1–C6 above are common conditions that are used in the literature. Similar conditions also were used in [28,30,39].

3.1. Asymptotic properties of the EL estimator

Theorem 1. *Under the conditions* C1–C3*, as $n\to\infty$, with probability tending to* 1 *the likelihood Eq.* (15) *has a solution $\tilde{\beta}$ within the open ball $\|\tilde{\beta}-\beta_0\|\le n^{-1/3}$.*

Theorem 1 provides the consistency of the MELE $\tilde{\beta}$. This result corresponds to Lemma 1 in [28] which is about the consistency of MELE for independent and identically distributed data. It is also an analogue of [4] for parametric maximum likelihood estimators.

Theorem 2. *Under the conditions* C1–C6,

$$\sqrt{n}(\tilde{\beta} - \beta_0) \longrightarrow N(0, V_{EL}) \tag{17}$$

in distribution as $n \to \infty$, where the covariance matrix V_{EL} is given by $V_{EL} = (\Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12})^{-1}$ and Σ_{12} and Σ_{11} have the same definition as (10).

Theorem 2 shows that the EL estimator $\tilde{\beta}$ is consistent and asymptotically normally distributed. It is interesting to note that the asymptotic covariance matrix of $\tilde{\beta}$ has the same form as that of the QIF estimator based on GMM. The asymptotic covariance matrix V_{EL} can be estimated consistently by

$$\left\{ \left[\sum_{i=1}^{n} \tilde{p}_i \frac{\partial S_i(\tilde{\beta})}{\partial \beta} \right]^T \left[\sum_{i=1}^{n} \tilde{p}_i S_i(\tilde{\beta}) S_i^T(\tilde{\beta}) \right]^{-1} \left[\sum_{i=1}^{n} \tilde{p}_i \frac{\partial S_i(\tilde{\beta})}{\partial \beta} \right] \right\}^{-1}$$

with $\tilde{p}_i = [n(1 + t^T(\tilde{\beta}) S_i(\tilde{\beta}))]^{-1}$, or alternatively by the same expression with $\tilde{p}_i$'s simply replaced by n^{-1} for simplicity.

It is noted that the proof of Theorem 1 is similar to the proof of Lemma 1 in [28], and the proof of Theorem 2 is similar to the proof of Theorem 1 in [28]. The details are thus omitted here but can be obtained from the authors on request.

3.2. Model checking and testing

Let us now turn our attention to the empirical log-likelihood ratio.

Theorem 3. *The empirical log-likelihood ratio statistic for testing $H_0 : \beta = \beta_0$ is*

$$W_E = 2\ell_n(\beta_0) - 2\ell_n(\tilde{\beta}). \tag{18}$$

Under the conditions C1–C6, *$W_E \to \chi_p^2$ in distribution as $n \to \infty$, when H_0 is true.*

Corollary 1. *Let $\beta = (\beta_1^T, \beta_2^T)^T$, where β_1 and β_2 are $r \times 1$ and $(p - r) \times 1$ vectors, respectively. For $H_0 : \beta_1 = \beta_1^0$, the profile empirical log-likelihood ratio test statistic is*

$$W_1 = 2\ell_n(\beta_1^0, \widetilde{\beta_2^0}) - 2\ell_n(\tilde{\beta}_1, \tilde{\beta}_2), \tag{19}$$

where $\widetilde{\beta_2^0}$ minimizes $\ell_n(\beta_1^0, \beta_2)$ with respect to β_2. Under the conditions C1–C6, *$W_2 \to \chi_r^2$ in distribution as $n \to \infty$, provided that H_0 is true.*

The proof of Theorem 3 and Corollary 1 can be achieved by using the similar arguments as those used in the proof of Theorem 2 and Corollary 5 in [28]. The details are thus omitted here but can be obtained upon request. It is noted that the QIF approach has also an analogue to the likelihood ratio test (ALRT), as shown by Theorem 1 in [30].

Clearly, Theorem 3 can be used not only to test the hypothesis $H_0 : \beta = \beta_0$ for some specific β_0, but also to construct the confidence region for β. In fact, let

$$I_\alpha = \{\beta : W_E \le c_\alpha\},$$

where c_α satisfies $Pr\{\chi_p^2 \le c_\alpha\} = \alpha$. Then, by Theorem 3 the set I_α gives a confidence region for the parameters β with asymptotically correct coverage probability $1 - \alpha$, i.e., $Pr(\beta \in I_\alpha) = 1 - \alpha + o_p(1)$. It is worth pointing out that the asymptotic confidence region I_α is obtained with no need to estimate the asymptotic covariance matrix of the resulting estimator of parameters.

4. Monte Carlo simulations

In this section, two simulation studies are made to assess the performance of the proposed approach in terms of the finite sample bias and the standard error of the parameter estimator, the coverage rate and average length of confidence intervals, as well as the power and size of hypothesis test.

4.1. Study 1: estimation of parameters

Consider the following model

$$y_{ij} = \beta x_{ij} + \epsilon_{ij}, \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, 10$$

Table 1
The empirical bias and standard deviation of β based on GEE, QIF and EL methods with different correlations and sample sizes, calculated from 1000 simulations. TR, true correlation structure; WR, working correlation structure.

			Bias			Standard deviation		
TR	WR	α	GEE	QIF	EL	GEE	QIF	EL
Sample size $n = 50$								
CS	AR1	0.3	−0.0003	−0.0005	−0.0005	0.0396	0.0389	0.0389
		0.5	−0.0004	−0.0005	−0.0004	0.0356	0.0352	0.0352
		0.7	−0.0004	−0.0003	−0.0002	0.0288	0.0289	0.0289
	CS	0.3	0.0004	0.0002	0.0004	0.0357	0.0353	0.0353
		0.5	0.0001	0.0001	0.0002	0.0308	0.0304	0.0305
		0.7	0.0000	0.0000	0.0003	0.0241	0.0238	0.0239
AR1	CS	0.3	0.0013	0.0010	0.0011	0.0394	0.0390	0.0390
		0.5	0.0013	0.0011	0.0012	0.0397	0.0393	0.0393
		0.7	0.0009	0.0008	0.0009	0.0378	0.0373	0.0373
	AR1	0.3	0.0005	0.0004	0.0004	0.0370	0.0366	0.0366
		0.5	0.0004	0.0003	0.0003	0.0330	0.0328	0.0328
		0.7	0.0002	0.0001	0.0001	0.0260	0.0265	0.0265
Sample size $n = 100$								
CS	AR1	0.3	−0.0001	−0.0001	−0.0002	0.0283	0.0280	0.0280
		0.5	−0.0002	−0.0001	−0.0001	0.0254	0.0254	0.0254
		0.7	−0.0001	0.0000	0.0000	0.0206	0.0208	0.0208
	CS	0.3	0.0004	0.0003	0.0004	0.0255	0.0253	0.0253
		0.5	0.0002	0.0002	0.0002	0.0219	0.0218	0.0218
		0.7	0.0001	0.0000	−0.0001	0.0172	0.0171	0.0171
AR1	CS	0.3	0.0007	0.0007	0.0007	0.0280	0.0279	0.0279
		0.5	0.0008	0.0007	0.0008	0.0282	0.0281	0.0281
		0.7	0.0006	0.0006	0.0007	0.0269	0.0267	0.0267
	AR1	0.3	0.0004	0.0004	0.0004	0.0264	0.0262	0.0262
		0.5	0.0002	0.0003	0.0003	0.0236	0.0236	0.0236
		0.7	0.0001	0.0002	0.0002	0.0185	0.0190	0.0190
Sample size $n = 200$								
CS	AR1	0.3	0.0006	0.0006	0.0006	0.0201	0.0199	0.0199
		0.5	0.0003	0.0004	0.0004	0.0181	0.0181	0.0181
		0.7	0.0001	0.0003	0.0003	0.0146	0.0148	0.0148
	CS	0.3	0.0007	0.0007	0.0007	0.0181	0.0180	0.0180
		0.5	0.0005	0.0005	0.0005	0.0156	0.0155	0.0155
		0.7	0.0004	0.0003	0.0000	0.0122	0.0122	0.0122
AR1	CS	0.3	0.0010	0.0011	0.0011	0.0199	0.0199	0.0199
		0.5	0.0010	0.0010	0.0011	0.0201	0.0200	0.0200
		0.7	0.0008	0.0009	0.0009	0.0191	0.0190	0.0190
	AR1	0.3	0.0008	0.0008	0.0008	0.0187	0.0187	0.0187
		0.5	0.0006	0.0007	0.0007	0.0167	0.0168	0.0168
		0.7	0.0003	0.0005	0.0005	0.0132	0.0136	0.0136

where $\beta = 1$, $x_i = (x_{i1}, \ldots, x_{i,10})^T$ is generated independently from a multivariate normal distribution with mean $(0.1, 0.2, \ldots, 1.0)^T$ and covariance matrix I where I is the identity matrix. And also $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i,10})^T$ is generated from a 10-dimensional normal distribution with mean 0, marginal variance 1 and the two common working correlation structures: compound symmetry (CS) and AR(1) with a single nuisance parameter α. In order to see how the proposed approach behaves under different degrees of correlation, the parameter α takes values 0.3, 0.5 and 0.7. In our simulation, we took sample size $n = 50$, 100 and 200, respectively. For each combination of n and α, we generated 1000 Monte Carlo random samples.

Table 1 reports the empirical bias and standard deviation of β based on the GEE, QIF and EL approaches from the 1000 simulation studies. Table 1 shows that the empirical biases are slightly smaller for QIF than EL in all but two cases. However, empirical biases from those three approaches are all smaller than 0.0013 and thus the biases can be neglectful. Hence, from Table 1 we conclude that those three methods all produce no detectable bias in the estimator of the regression coefficient β, which confirms the theoretical result. As sample size increases, the standard deviation for the estimator of β by those three approaches decreases. An interesting finding observed from Table 1 is that there is no difference between the estimated standard deviations by the QIF and EL approaches. This is reasonable as we already showed that the resulting estimator from the EL approach has the same asymptotic covariance form as that of the QIF estimator. In fact, in those two methods we use the same formula to estimate the standard deviation of β except for using different consistent estimators of β to replace β in the formula.

Next, we examine the performance for each approach in terms of the coverage probability and average length of the resulting confidence interval. Table 2 presents the empirical coverage probability and average length of the confidence interval over 1000 simulation studies, based on the GEE, QIF and EL approaches, respectively. For the QIF approach, we constructed the confidence interval using two different methods, say QIF1 and QIF2. In the QIF1, the confidence interval is

Table 2
The average length and empirical coverage rates for confidence interval of β based on GEE, QIF and EL methods with different correlations and sample sizes, calculated from 1000 simulations. TR, true correlation structure; WR, working correlation structure; QIF1, QIF confidence interval based on the asymptotic normality; QIF2, QIF confidence interval based on the ALRT test.

			Average length				Coverage rate			
TR	WR	α	GEE	QIF1	QIF2	EL	GEE	QIF1	QIF2	EL
Sample size $n = 50$										
CS	AR1	0.3	0.1552	0.1525	0.1349	0.1275	0.937	0.933	0.952	0.933
		0.5	0.1395	0.1381	0.1218	0.1152	0.939	0.931	0.947	0.940
		0.7	0.1130	0.1133	0.0998	0.0943	0.939	0.935	0.953	0.942
	CS	0.3	0.1400	0.1385	0.1222	0.1162	0.938	0.936	0.947	0.939
		0.5	0.1206	0.1194	0.1052	0.1001	0.941	0.940	0.949	0.936
		0.7	0.0944	0.0934	0.0822	0.0783	0.944	0.940	0.952	0.934
AR1	CS	0.3	0.1544	0.1527	0.1353	0.1283	0.940	0.933	0.942	0.940
		0.5	0.1556	0.1539	0.1360	0.1294	0.944	0.939	0.941	0.937
		0.7	0.1481	0.1463	0.1290	0.1237	0.947	0.940	0.941	0.934
	AR1	0.3	0.1449	0.1434	0.1268	0.1202	0.936	0.936	0.949	0.936
		0.5	0.1293	0.1287	0.1136	0.1080	0.936	0.939	0.940	0.933
		0.7	0.1018	0.1037	0.0915	0.0870	0.942	0.943	0.940	0.935
Sample size $n = 100$										
CS	AR1	0.3	0.1109	0.1096	0.0948	0.0926	0.936	0.940	0.955	0.952
		0.5	0.0997	0.0994	0.0861	0.0840	0.936	0.938	0.959	0.954
		0.7	0.0808	0.0816	0.0704	0.0686	0.940	0.938	0.957	0.951
	CS	0.3	0.0998	0.0993	0.0858	0.0839	0.941	0.940	0.957	0.955
		0.5	0.0860	0.0856	0.0737	0.0721	0.942	0.934	0.958	0.947
		0.7	0.0673	0.0670	0.0576	0.0563	0.942	0.938	0.959	0.952
AR1	CS	0.3	0.1098	0.1093	0.0946	0.0924	0.943	0.944	0.959	0.953
		0.5	0.1107	0.1101	0.0953	0.0932	0.937	0.936	0.957	0.954
		0.7	0.1053	0.1047	0.0903	0.0890	0.936	0.933	0.959	0.953
	AR1	0.3	0.1033	0.1028	0.0878	0.0858	0.939	0.938	0.949	0.945
		0.5	0.0923	0.0924	0.0788	0.0769	0.939	0.940	0.952	0.944
		0.7	0.0727	0.0746	0.0639	0.0624	0.950	0.942	0.951	0.949
Sample size $n = 200$										
CS	AR1	0.3	0.0788	0.0781	0.0669	0.0662	0.945	0.942	0.955	0.952
		0.5	0.0708	0.0708	0.0608	0.0603	0.950	0.944	0.956	0.956
		0.7	0.0574	0.0581	0.0500	0.0495	0.941	0.943	0.954	0.952
	CS	0.3	0.0709	0.0707	0.0604	0.0598	0.954	0.949	0.952	0.950
		0.5	0.0611	0.0609	0.0522	0.0517	0.952	0.950	0.953	0.950
		0.7	0.0478	0.0477	0.0409	0.0409	0.953	0.948	0.952	0.946
AR1	CS	0.3	0.0781	0.0779	0.0664	0.0657	0.947	0.949	0.957	0.952
		0.5	0.0786	0.0784	0.0670	0.0663	0.940	0.949	0.958	0.959
		0.7	0.0747	0.0745	0.0635	0.0630	0.941	0.939	0.955	0.952
	AR1	0.3	0.0734	0.0733	0.0621	0.0615	0.947	0.948	0.948	0.947
		0.5	0.0656	0.0659	0.0560	0.0554	0.951	0.941	0.952	0.951
		0.7	0.0517	0.0531	0.0454	0.0450	0.948	0.947	0.960	0.956

constructed based on the asymptotic normality distribution, while in the QIF2 the confidence interval is formed using the ALRT test, as shown in Theorem 1 in [30].

The simulation findings in Table 2 can be summarized as follows. First of all, as the sample size increases, the average length of confidence interval given by those approaches decreases and the corresponding empirical coverage probability is more close to the nominal coverage probability 0.95. When the sample size is small, say $n = 50$, the QIF2 approach based on the ALRT test, produces the highest empirical coverage probabilities for most cases. The GEE approach gives higher coverage probabilities than the QIF1 and EL methods, although the corresponding empirical coverage probabilities are smaller than the nominal coverage probability 0.95. When the sample size is moderate or large, for example, $n = 100$ or 200, our simulation shows that compared to the GEE and QIF approaches, the empirical coverage probabilities for the confidence intervals given by the EL approach are more close to the nominal coverage probability 0.95.

Second, correct specification of the correlation structure yields a short average length of confidence interval obtained by each of those approaches.

Third, for all the simulation cases the EL-based confidence interval lengths are shorter than those by the GEE, QIF1 and QIF2, although the average length of the EL-based confidence interval is very close to those by the QIF2, especially when the sample size is moderate or large. Both the GEE and QIF1 approaches give longer confidence intervals than the EL approach for all the simulation cases.

Overall, the proposed EL approach has better performance than the GEE approach in terms of coverage probability and the average length of confidence interval. Although the EL approach gives the same asymptotic covariance matrix for the resulting estimator for the regression coefficients as the QIF method, the EL approach has better finite-sample performance than the QIF method.

Table 3
Simulated power for testing $H_0 : \beta_2 = \beta_2^0$ by GEE, QIF and EL methods in the simulation Study 3 with the true correlation structure CS and $\rho = 0.7$. QIF.Wald is the Wald test based on the asymptotic normality of the QIF estimator and QIF.ALRT is the quadratic score test based on a quadratic inference function.

Working R	β_2^0	GEE.Wald	QIF.Wald	QIF.ALRT	EL.ELR
AR(1)	1	0.059	0.048	0.039	0.048
	1.05	0.381	0.418	0.386	0.413
	1.1	0.887	0.920	0.900	0.914
CS	1	0.052	0.050	0.037	0.051
	1.05	0.512	0.530	0.486	0.532
	1.1	0.957	0.976	0.964	0.972

4.2. Study 2: test of hypothesis

As suggested by an anonymous reviewer, in this section we conduct Monte Carlo simulation experiments to investigate the small sample behavior of the empirical log-likelihood ratio statistic presented in Theorem 3. We consider the following model

$$y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \epsilon_{ij}, \quad i = 1, \dots, n; j = 1, 2, \dots, m,$$

where $\beta_1 = 0.5, \beta_2 = 1, n = 100$, and $m = 5$. x_{ij1} are i.i.d. from Uniform(0.5, 1.5), and x_{ij2} are independently from $N(\mu_{x_{ij2}}, 1)$ with $\mu_{x_{ij2}} = j/m$ for $j = 1, 2, \dots, m$ and $i = 1, 2 \dots, n$. The random errors $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{im}) \sim N_m(0, R(\rho))$ where the covariance matrix $R(\rho)$ is of the compound symmetry (CS) structure with $\rho = 0.7$. We aim to test the null hypothesis $H_0 : \beta_2 = \beta_2^0$, where β_2^0 takes different values at $\beta_2^0 = 1.00$, 1.05 and 1.10 in order to assess the test size and power.

For the GEE approach, based on the asymptotic normality of regression parameter estimators we can construct the Wald test for the null hypothesis $H_0 : \beta_2 = \beta_2^0$, denoted by GEE.Wald. For the QIF approach, there are two versions of testing statistic for the above null hypothesis H_0. One is similar to the test method GEE.Wald, which is using the Wald test principle and the asymptotic normality distribution of the QIF regression parameter estimator, denoted by QIF.Wald. The other is constructed using an analogue of the likelihood ratio test (ALRT) for the QIF, denoted by QIF.ALRT. For the EL approach, the empirical log-likelihood ratio (ELR) statistic presented in Theorem 3 is applied, denoted by EL.ELR.

Table 3 displays the simulation results based on 1000 replicates for the four aforementioned test methods. It shows that both the QIF.Wald test and the EL.ELR test perform very well, in particular when the within-subject correlation structure is correctly specified. Those two test methods behave quite similarly in terms of test size and power. When the working correlation structure is wrongly specified as AR(1), the GEE.Wald test has slightly inflated type I error and the size of QIF.ALRT based test is smaller than the nominal size. When the working correlation structure is correctly specified as CS, the test sizes of all tests except for QIF.ALRT test are very closed to the nominal size. On the other hand, the QIF.Wald test appears most powerful, while the GEE.Wald test is least powerful in most cases regardless of whether or not the working correlation structure is correctly specified.

5. Analysis of Ohio Children's wheeze status data

We apply the proposed method to analyze Ohio Children's wheeze status data, which is part of the longitudinal binary data on respiratory health effects of indoor and outdoor air pollution in six U.S. cities. One of the interests of the study is to determine the effect of maternal smoking on children's respiratory illness. This data set was analyzed by many authors [41,8,30]. They analyzed the data on 537 children collected on at ages 7 to 10 from Ohio and treated the maternal smoking habit as fixed at the first visit.

The response is a binary outcome with 0 and 1, indicating the presence or absence of respiratory illness. The maternal smoking habit, in the preceding year, is recorded as a binary covariate. For further details about the data background, see [8]. One of the aims of the study was to assess the effect of maternal smoking on children's respiratory illness. Clearly, we can expect that the measurements for the same child are very likely serially correlated.

The following logistic model is applied to the binary data

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1} x_{ij2}$$

for $i = 1, \dots, 537$ and $j = 1, \dots, 4$, where x_{ij1}, x_{ij2} and $x_{ij1}x_{ij2}$ are the age of the child, the maternal smoking habit indicator and their interaction, respectively. The matrix A_i is diagonal with elements $v(\pi_{it}) = \pi_{it}(1 - \pi_{it})$. The extended score vector $\bar{S}_n(\beta)$ is constructed by choosing $k = 2$, for instance, by choosing the working correlation structure as either CS or AR(1) structure, so that only the first two basis matrices are used in this case.

Table 4 presents the point estimators and their corresponding standard errors, obtained by the GEE, QIF and EL methods under the two commonly used working correlations CS and AR(1). The estimators of the regression parameters are very similar for the three methods under the CS and AR(1) working correlation structures. When the working correlation structure is specified as the CS, the standard errors for β_2 and β_3 by GEE are bit smaller than those by the QIF and EL. When the

Table 4
The parameter estimators and corresponding standard errors (in the parentheses) for Ohio Children's wheeze status data by GEE, QIF and EL methods.

Working R	Parameter	GEE	QIF	EL
CS	β_0	−1.9005(0.1191)	−1.9120(0.1194)	−1.9008(0.1195)
	β_1	−0.1412(0.0582)	−0.1456(0.0585)	−0.1413(0.0586)
	β_2	0.3138(0.1878)	0.2879(0.2124)	0.3140(0.2012)
	β_3	0.0708(0.0883)	0.0760(0.0905)	0.0708(0.0902)
AR(1)	β_0	−1.9195(0.1200)	−1.9170(0.1198)	−1.9071(0.1197)
	β_1	−0.1468(0.0593)	−0.1469(0.0586)	−0.1444(0.0589)
	β_2	0.2953(0.1900)	0.2868(0.1902)	0.3142(0.1903)
	β_3	0.0815(0.0907)	0.0783(0.0900)	0.0757(0.0905)

Table 5
Testing of hypotheses for the longitudinal data on Children's wheeze status. The column $2\ell_n(\tilde{\beta})$ is minimum of the log-likelihood obtained under the null hypotheses, W_1 is the value of the test statistic, df is the degrees of freedom.

Null hypotheses	$2\ell_n(\tilde{\beta})$	W_1	df	p-value
Full model	3.972	0	–	–
$H_0 : \beta_4 = 0$	4.449	0.477	1	0.490
$H_0 : \beta_2 = \beta_3 = 0$	5.823	1.851	2	0.396
$H_0 : \beta_1 = \beta_3 = 0$	10.337	6.365	2	0.041
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$	11.898	7.926	3	0.048

working correlation structure is specified as the AR(1), the standard errors for β_1 and β_3 by both the QIF and EL are smaller than those by the GEE. This difference may be due to the fact that the true correlation structure is unknown. In fact, the approximation of the inverse of the working matrix, R, used in both the QIF and EL approaches, depends on the choice of the working correlation structure. The t-ratio shows that the interaction between the age of the child and maternal smoking is not significant. The t-ratio also suggests that age is a significant factor and should be included in the model. The negative sign for age means that older children are less likely to have respiratory disease. Similarly, the maternal smoking habit has a positive effect on children's respiratory disease, although smoking habit seems not to be a significant factor.

In order to assess whether the sub-models are adequate, we compute the statistic $2\ell_n(\tilde{\beta})$ under various sub-models with certain parameter constraints and perform the ELR tests to compare different models. Results are reported in Table 5.

Each row of Table 5 represents results for a given model with the specific constraints on the parameters. In particular, the first row provides the full model that includes all the covariates, age, maternal smoking and their interaction. For each model, we compute the parameter estimate $\tilde{\beta}$ and report the corresponding value of $2\ell_n(\tilde{\beta})$. The test statistic W_1 is calculated using (19) which is actually compared to the full model. The column of "df" is the degrees of freedom for the corresponding test statistic and the last column gives the associated p-value. From Table 5, it is clear that the null hypotheses $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ and $H_0 : \beta_1 = \beta_3 = 0$ are rejected as the corresponding p-values are smaller than 0.05. In contrast, the remaining null hypotheses $H_0 : \beta_2 = \beta_3 = 0$ and $H_0 : \beta_4 = 0$ cannot be rejected.

6. Conclusions

The EL approach is a powerful tool for combining information about parameters when there are more estimating equations than those actually needed for estimation of the unknown parameters. In this paper, we proposed EL-based inference for longitudinal data within the framework of generalized linear models. The proposed approach takes into account the within-subject correlation but has no need to estimate the nuisance parameters involved in the correlation matrix. The resulting estimator of parameters retains optimal even if the working correlation structure is misspecified. The proposed approach yields more efficient estimators than the conventional GEE approach and achieves the same asymptotic covariance as the QIF method. Compared to the QIF method, the proposed EL approach has a better finite-sample performance in the sense of the average length of confidence interval and coverage probability of the estimator of regression coefficients.

It is noted that Leung et al. [15] recently proposed a hybrid GEE method for longitudinal data by using the EL approach to combine multiple GEEs based on different working correlation models. The EL approach is employed in both our method and Leung et al. [15]'s approach, but it is used in very different ways. In fact, in our EL method the estimating equations are different from those by Leung et al. [15]. More specifically, Leung et al. [15] retained the idea of GEE and focused on the manner of combining GEEs which are based on different working correlation structures. In their method there are more than one working correlation matrices and each working correlation matrix is used to form one GEE. In our proposed approach, however, we proposed to use only one working correlation matrix whose inverse is approximated by a linear combination of basis matrices. We then combine the resulting "extended score" vector in terms of the EL method. Therefore, our proposed method is simpler than Leung et al. [15]'s one in this sense. More importantly, in Leung et al. [15]'s approach the nuisance parameters in the working correlation matrices have to be estimated, which is not always feasible even for some simple cases [5]. In contrast, there is no need to estimate any nuisance parameters in our method.

Acknowledgments

Li's research was supported by an Overseas Research Scholarship from the UK and a Scholarship from School of Mathematics, University of Manchester, UK. Pan's research was supported by a grant from the Royal Society of the UK. We would like to thank two anonymous reviewers for their constructive comments and helpful suggestions. We are also grateful to Professor Annie Qu for providing the MATLAB code for the QIF method.

References

[1] Y. Bai, W. Fung, Z. Zhu, Penalized quadratic inference functions for single-index models with longitudinal data, Journal of Multivariate Analysis 100 (2009) 152–161.
[2] Y. Bai, W. Fung, Z. Zhu, Weighted empirical likelihood for generalized linear models with longitudinal data, Journal of Statistical Planning and Inference 140 (2010) 3446–3456.
[3] Y. Bai, Z. Zhu, W. Fung, Partial linear models for longitudinal data based on quadratic inference functions, Scandinavian Journal of Statistics 35 (2008) 104–118.
[4] H. Cramér, Mathematical Methods of Statistics, Princeton University Press, 1946.
[5] M. Crowder, On the use of a working correlation matrix in using generalised linear models for repeated measures, Biometrika 82 (1995) 407–410.
[6] T. DiCiccio, P. Hall, J. Romano, Empirical likelihood is Bartlett-correctable, The Annals of Statistics 19 (1991) 1053–1061.
[7] P.J. Diggle, P.J. Heagerty, K.Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, 2002.
[8] G. Fitzmaurice, N. Laird, A likelihood-based method for analysing longitudinal binary responses, Biometrika 80 (1993) 141–151.
[9] P. Hall, B. La Scala, Methodology and algorithms of empirical likelihood, International Statistical Review 58 (1990) 109–127.
[10] L. Hansen, Large sample properties of generalized method of moments estimators, Econometrica 50 (1982) 1029–1054.
[11] G. Imbens, One-step estimators for over-identified generalized method of moments models, The Review of Economic Studies 64 (1997) 359.
[12] G. Imbens, Generalized method of moments and empirical likelihood, Journal of Business and Economic Statistics 20 (2002) 493–506.
[13] E. Kolaczyk, Empirical likelihood for generalized linear models, Statistica Sinica 4 (1994) 199–218.
[14] S. Lele, M. Taper, A composite likelihood approach to (co)variance components estimation, Journal of Statistical Planning and Inference 103 (2002) 117–135.
[15] D. Leung, Y. Wang, M. Zhu, Efficient parameter estimation in longitudinal data analysis using a hybrid GEE method, Biostatistics 10 (2009) 436–445.
[16] H. Liang, Y. Qin, X. Zhang, D. Ruppert, Empirical likelihood-based inferences for generalized partially linear models, Scandinavian Journal of Statistics 36 (2009) 433–443.
[17] K. Liang, S. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[18] G. Li, L. Zhu, L. Xue, S. Feng, Empirical likelihood inference in partially linear single-index models for longitudinal data, Journal of Multivariate Analysis 101 (2010) 718–732.
[19] P. McCullagh, Quasi-likelihood functions, The Annals of Statistics 11 (1983) 59–67.
[20] P. McCullagh, J. Nelder, Generalized Linear Models, Chapman & Hall/CRC, 1989.
[21] J. Nelder, R. Wedderburn, Generalized linear models, Journal of the Royal Statistical Society. Series A (General) 135 (1972) 370–384.
[22] W. Newey, R. Smith, Higher order properties of gmm and generalized empirical likelihood estimators, Econometrica 72 (2004) 219–255.
[23] A. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[24] A. Owen, Empirical likelihood ratio confidence regions, The Annals of Statistics 18 (1990) 90–120.
[25] A. Owen, Empirical likelihood for linear models, The Annals of Statistics 19 (1991) 1725–1747.
[26] A. Owen, Empirical Likelihood, Chapman & Hall/CRC Press, 2001.
[27] F. Pukelsheim, Optimal Design of Experiments, Wiley, New York, NY, 1993.
[28] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, The Annals of Statistics 22 (1994) 300–325.
[29] A. Qu, R. Li, Quadratic inference functions for varying-coefficient models with longitudinal data, Biometrics 62 (2006) 379–391.
[30] A. Qu, B. Lindsay, B. Li, Improving generalised estimating equations using quadratic inference functions, Biometrika 87 (2000) 823–836.
[31] J. Shi, T. Lau, Empirical likelihood for partially linear models, Journal of Multivariate Analysis 72 (2000) 132–148.
[32] Y. Wang, V. Carey, Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance, Biometrika 90 (2003) 29–41.
[33] Q. Wang, B. Jing, Empirical likelihood for partial linear models, Annals of the Institute of Statistical Mathematics 55 (2003) 585–595.
[34] S. Wang, L. Qian, R. Carroll, Generalized empirical likelihood methods for analyzing longitudinal data, Biometrika 97 (2010) 79–93.
[35] R. Wedderburn, Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, Biometrika 61 (1974) 439–447.
[36] L. Xue, L. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, Journal of the American Statistical Association 102 (2007) 642–654.
[37] L. Xue, L. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, Biometrika 94 (2007) 921–937.
[38] H. Yang, T. Li, Empirical likelihood for semiparametric varying coefficient partially linear models with longitudinal data, Statistics & Probability Letters 80 (2010) 111–121.
[39] H. Ye, J. Pan, Modelling of covariance structures in generalised estimating equations for longitudinal data, Biometrika 93 (2006) 927–941.
[40] J. You, G. Chen, Y. Zhou, Block empirical likelihood for longitudinal partially linear regression models, Canadian Journal of Statistics 34 (2006) 79–96.
[41] S. Zeger, K. Liang, P. Albert, Models for longitudinal data: a generalized estimating equation approach, Biometrics 44 (1988) 1049–1060.
[42] Y. Zhao, W. Jian, Analysis of longitudinal data in the case-control studies via empirical likelihood, Communications in Statistics: Simulation and Computation 36 (2007) 565–578.
[43] P. Zhao, L. Xue, Empirical likelihood inferences for semiparametric varying-coefficient partially linear errors-in-variables models with longitudinal data, Journal of Nonparametric Statistics 21 (2009) 907–923.
[44] P. Zhao, L. Xue, Empirical likelihood inferences for semiparametric varying-coefficient partially linear models with longitudinal data, Communications in Statistics. Theory and Methods 39 (2010) 1898–1914.