

Generalized empirical likelihood methods for analyzing longitudinal data

BY SUOJIN WANG

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
sjwang@stat.tamu.edu

LIANFEN QIAN

Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, Florida 33431, U.S.A.
lqian@fau.edu

AND RAYMOND J. CARROLL

Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
carroll@stat.tamu.edu

SUMMARY

Efficient estimation of parameters is a major objective in analyzing longitudinal data. We propose two generalized empirical likelihood based methods that take into consideration within-subject correlations. A nonparametric version of the Wilks theorem for the limiting distributions of the empirical likelihood ratios is derived. It is shown that one of the proposed methods is locally efficient among a class of within-subject variance-covariance matrices. A simulation study is conducted to investigate the finite sample properties of the proposed methods and compare them with the block empirical likelihood method by You et al. (2006) and the normal approximation with a correctly estimated variance-covariance. The results suggest that the proposed methods are generally more efficient than existing methods which ignore the correlation structure, and better in coverage compared to the normal approximation with correctly specified within-subject correlation. An application illustrating our methods and supporting the simulation study results is also presented.

Some key words: Confidence region; Efficient estimation; Empirical likelihood; Longitudinal data; Maximum empirical likelihood estimator.

1. INTRODUCTION

Empirical likelihood as a nonparametric data-driven technique was first proposed by Owen (1988). It employs the likelihood function without specifically assuming a distribution for the data and incorporates side information, through constraints or a prior distribution, which maximizes the efficiency of the method.

Longitudinal data are often collected from experimental studies or clinical trials. The subjects under study are measured repeatedly during the period of the study. A major part 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 one would expect that it

would lead to a loss of efficiency for empirical likelihood applications. The major purpose of this paper is to demonstrate this fact and show how to exploit the within-subject correlation.

In the past two decades, unique and desirable properties of the empirical likelihood method have been studied by a number of authors. These properties include, but are not limited to, range respecting, transformation-preserving, asymmetric confidence intervals, Bartlett correctability and better coverage probability or shorter confidence length for non-longitudinal data when the sample size is small to moderate. Empirical likelihood has been used to estimate model parameters in situations such as linear models, possibly with missing data, generalized linear models and partial linear models. Owen (2001) provides a comprehensive account of empirical likelihood and its properties. More recently, this methodology has been extended to other problems, such as mixture models in case-control studies (Zou et al., 2002) and analyzing censored or survival data (Li & Wang, 2003; Zhao, 2005).

You et al. (2006) constructed a subject-wise empirical likelihood method for longitudinal data, using working independence and thus ignoring the within-subject correlation structure. Later, Xue & Zhu (2007) proposed another subject-wise empirical likelihood method through centering longitudinal data and obtained asymptotic normality of the maximum empirical likelihood estimator of the regression coefficients and a nonparametric version of the Wilks theorem. Xue & Zhu (2007) do not consider the within-subject correlation structure of the longitudinal data. Both You et al. (2006) and Xue & Zhu (2007) address partial linear models.

In this paper we show how to incorporate the within-subject correlation structure of the repeated measurements into the empirical likelihood method. We also show how our methods can be used to perform empirical likelihood inference on individual components of the regression function. We will consider the following continuous response variable longitudinal regression model

$$y_{ij} = X_{ij}^T \beta + \epsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, m_i, \quad (1)$$

where n is the number of subjects participating in the study, m_i is the number of repeated measurements for the i th subject, y_{ij} is the j th measurement on the i th subject, X_{ij} is a q -vector of covariate values, β is a q -vector of unknown regression coefficients and ϵ_{ij} is a zero-mean random variable with variance σ_{ij}^2 representing the deviation of the response from the model prediction, $X_{ij}^T \beta$. In matrix notation, let $Y_i^T = (y_{i1}, \dots, y_{im_i})$, $X_i^T = (X_{i1}, \dots, X_{im_i})$, $X_{ij}^T = (x_{ij1}, \dots, x_{ijq})$, $\beta^T = (\beta_1, \dots, \beta_q)$, $\epsilon_i^T = (\epsilon_{i1}, \dots, \epsilon_{im_i})$ and $N = \sum_{i=1}^n m_i$. Then model (1) can be rewritten as

$$Y_i = X_i \beta + \epsilon_i. \quad (2)$$

In (2), the elements within each ϵ_i are allowed to be correlated. Let

$$Z_i(\beta) = \{z_{i1}(\beta), \dots, z_{im_i}(\beta)\}^T = V_i^{*-1}(Y_i - X_i \beta), \quad (3)$$

where V_i^* is an $m_i \times m_i$ invertible matrix. If $V_i^* = V_i = \text{cov}(Y_i | X_i)$ for all $i = 1, \dots, n$, then $Z_i(\beta)$ is the usual generalized least squares estimating function. If $V_i^* = \sigma^2 I_{m_i}$ for some $\sigma^2 > 0$, where I_{m_i} is the m_i dimensional identity matrix, then $Z_i(\beta)$ is the estimating function with the working independence assumption. As will be shown in this paper, for empirical likelihood $V_i^* = V_i$ for all i is desired by an optimality measure.

More generally, we can define an estimating equation

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta) = 0, \quad (4)$$

where $\Phi_{ij}(\beta)$ is the estimating function. In (2)-(3), $\Phi_{ij}(\beta) = X_{ij}z_{ij}(\beta)$. In this paper, we propose two generalized empirical likelihood methods to accommodate the correlation structure in model (1). A nonparametric version of the Wilks theorem is derived. We will also address inference for linear combinations of the regression parameters for both proposed methods. Further, we will show that the subject-wise test is locally efficient and demonstrate the efficiency of our methods in simulation and an empirical example.

2. TWO NEW EMPIRICAL LIKELIHOOD ESTIMATORS

2.1. Definitions

Within the empirical likelihood context, You et al. (2006) suggested a subject-wise empirical likelihood based estimation for regression coefficients using working independence correlation matrices, a method equivalent to treating the correlation matrices as the identity. It is well known that the working independence assumption may lead to a loss of efficiency in estimation when within-subject correlation exists.

In this section we propose two ways to define an empirical likelihood, namely

- (I). Element-wise empirical likelihood: assigning a probability mass p_{ij} to each observation y_{ij} ;
- (II). Subject-wise empirical likelihood: assigning a probability p_i to each subject i as is done in You et al. (2006).

In the current literature on empirical likelihood methods, the potential correlation structure of within-subject observations has not been taken into consideration. Here we demonstrate that in each of the two approaches above, it is easy to incorporate the correlation structure and the resulting empirical likelihoods are simple to implement.

2.2. Element-wise Empirical Likelihood Approach

Here we propose a new approach for element-wise empirical likelihood. Let β_0 be the true parameter vector satisfying model (1). Let \widehat{V}_{in} be an estimator of V_i . How to estimate V_i effectively is a challenging problem, see for example Ye & Pan (2006) and the references therein. To take into account possible inconsistency in the covariance matrix estimation typically caused by a misspecified correlation structure, throughout this paper we assume only that \widehat{V}_{in} converges to some V_i^* given in (3) in probability uniformly over all $i = 1, \dots, n$. Such convergence for each subject i is generally possible under certain regularity conditions in generalized estimating equations, e.g., when we assume a common parametric model for V_i across all i and the total number of unknown parameters is bounded. The generalized estimating equations setup is generally based upon this assumption. Conditions like this can be weakened further at the expense of more delicate treatments by assuming that the convergence for some summary statistics over all subjects in approximating equations such as (A2) in the Appendix is valid, but we will not consider this further treatment in this paper.

In our numerical illustrations, we use the following simple methods: (a) the nonparametric sample covariance matrix obtained from the residuals of a generalized estimating equation fit using working independence and (b) the maximum likelihood estimator under an assumed specification of the correlation structure. When the correlation structure is correctly specified then the estimator \widehat{V}_{in} is a consistent estimator of $V_i^* = V_i$. We emphasize however that our asymptotic results only require that the \widehat{V}_{in} converge uniformly to some V_i^* .

Let $N = \sum_{i=1}^n m_i$ and define $\widehat{Z}_i(\beta) = \{\widehat{z}_{i1}(\beta), \dots, \widehat{z}_{im_i}(\beta)\}^T = \widehat{V}_{in}^{-1}(Y_i - X_i\beta)$ and $\widehat{\Phi}_{ij}(\beta) = X_{ij}\widehat{z}_{ij}(\beta)$. We define the element-wise empirical likelihood ratio function as

$$L_1(\beta) = \sup_{p_{ij}} \left\{ \prod_{i=1}^n \prod_{j=1}^{m_i} N p_{ij} \mid p_{ij} \geq 0, \sum_{i=1}^n \sum_{j=1}^{m_i} p_{ij} = 1, \sum_{i=1}^n \sum_{j=1}^{m_i} p_{ij} \widehat{\Phi}_{ij}(\beta) = 0 \right\}.$$

Let $\widehat{G}_{in} = \widehat{V}_{in}^{-1}$ with elements \widehat{g}_{ijk} . Then by assumption $\widehat{G}_{in} = V_i^{*-1} + o_p(1)$ uniformly. Denote the elements of V_i^{*-1} by g_{ijk} . Before we state one of the main results, we note the following regularity conditions.

Condition 1. As $n \rightarrow \infty$, $P[0 \in ch\{\Phi_{ij}(\beta_0), i = 1, \dots, n; j = 1, \dots, m_i\}] \rightarrow 1$, where $ch\{\}$ is the convex hull.

Condition 2. The limits $\Sigma_1 = \lim_{n \rightarrow \infty} N^{-1} \sum_{i=1}^n X_i^T V_i^{*-1} V_i V_i^{*-1} X_i$ and $\Sigma_2 = \lim_{n \rightarrow \infty} N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} X_{ij} d_i(j) X_{ij}^T$ exist and are positive definite, where $d_i(j)$ is the j th diagonal element of $V_i^{*-1} V_i V_i^{*-1}$.

Condition 3. $E\|\Phi_{ij}(\beta_0)\|^{2+\eta}$ are bounded uniformly for some $\eta > 0$.

Condition 4. For each pair i and i' with $1 \leq i, i' \leq n$ and $i \neq i'$, $\widehat{g}_{ijk} - \widehat{g}_{-(i,i')jk} = O_p(n^{-1})$ and sufficient moment conditions are satisfied so that $E(\widehat{B}_{ii'}) = O(n^{-1})$ and $E(\widehat{B}_{ii'} \widehat{B}_{ii'}^T) = O(n^{-2})$, where $\widehat{g}_{-(i,i')jk}$ is \widehat{g}_{ijk} but computed with all the data except for subjects i and i' and $\widehat{B}_{ii'} = \sum_{j=1}^{m_i} \sum_{k=1}^{m_{i'}} (\widehat{g}_{ijk} - \widehat{g}_{-(i,i')jk}) X_{ij} \epsilon_{ik}$.

In addition, assume that $\max_{1 \leq i \leq n} (m_i)$ is bounded for all n . It is worth mentioning that Condition 4 can be weakened at the cost of considerably more notational complexity. Under Conditions 1-4, in the Appendix we show the following result.

THEOREM 1. *Let Σ_1 and Σ_2 be the two matrices defined in Condition 2. Then as $n \rightarrow \infty$,*

$$-2 \log\{L_1(\beta_0)\} \rightarrow \sum_{k=1}^q c_k \chi_{1k}^2$$

in distribution, where c_k are the eigenvalues of $\Sigma_1^{1/2} \Sigma_2^{-1} \Sigma_1^{1/2}$, or equivalently of $\Sigma_2^{-1} \Sigma_1$, and χ_{1k}^2 are independent χ^2 random variables.

In Theorem 1 it is not required that $V_i^* = V_i$. However, if \widehat{V}_{in} is a consistent estimator of V_i , then $V_i^{*-1} V_i V_i^{*-1} = V_i^{-1}$ in the definitions of Σ_1 and Σ_2 in Condition 2. Moreover, in the special case of $V_i^* = V_i = \sigma^2 I_{m_i}$ when the within subject responses are uncorrelated, all $c_k = 1$ and thus $-2 \log\{L_1(\beta_0)\} \rightarrow \chi_q^2$ in distribution.

Let $\Sigma_1(\beta)$ be the asymptotic covariance matrix of $N^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta)$ and $\widehat{\Sigma}_{1N}(\beta)$ be a consistent estimator of $\Sigma_1(\beta)$, such as $N^{-1} \sum_{i=1}^n \widehat{\Phi}_i(\beta) \widehat{\Phi}_i^T(\beta)$, where $\widehat{\Phi}_i(\beta) = \sum_{j=1}^{m_i} \widehat{\Phi}_{ij}(\beta)$ for $\Phi_i(\beta) = \sum_{j=1}^{m_i} \Phi_{ij}(\beta)$. Define

$$\widehat{\Sigma}_{2N}(\beta) = N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{\Phi}_{ij}(\beta) \widehat{\Phi}_{ij}^T(\beta),$$

$$S_N(\beta) = N^{-1} \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{\Phi}_{ij}(\beta) \right\} \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{\Phi}_{ij}(\beta) \right\}^T.$$

Then following Rao & Scott (1981), we consider the adjustment factor

$$a_N(\beta) = \text{tr} \left\{ \widehat{\Sigma}_{1N}^{-1}(\beta) S_N(\beta) \right\} / \text{tr} \left\{ \widehat{\Sigma}_{2N}^{-1}(\beta) S_N(\beta) \right\},$$

where “tr” stands for “trace”.

THEOREM 2. *If Conditions 1-4 hold, then as $n \rightarrow \infty$,*

$$-2a_N(\beta_0) \log\{L_1(\beta_0)\} \rightarrow \chi_q^2$$

in distribution.

Proof. Denote $l_1(\beta_0) = -2 \log\{L_1(\beta_0)\}$. Notice that $l_1(\beta_0)$ is an asymptotically quadratic form of a matrix, with the sandwich matrix $\widehat{\Sigma}_{2N}^{-1}(\beta_0)$, see (A2) in the Appendix, and hence equals $\text{tr}\{\widehat{\Sigma}_{2N}^{-1}(\beta_0) S_N(\beta_0)\} + o_p(1)$. The proof is completed by noting that $\text{tr}\{\widehat{\Sigma}_{1N}^{-1}(\beta_0) S_N(\beta_0)\} \rightarrow \chi_q^2$ in distribution.

2.3. Subject-wise Empirical Likelihood Method

We now propose another empirical likelihood based on subject-wise units. Define the subject-wise empirical likelihood ratio as

$$L_2(\beta) = \sup_{p_i} \left\{ \prod_{i=1}^n n p_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \widehat{\Phi}_i(\beta) = 0 \right\}.$$

We assume regularity conditions similar to Conditions 1-4, namely

Condition 1'. Condition 1 holds when applied to $\Phi_i(\beta_0)$.

Condition 2'. The limit $\Sigma_3 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T V_i^*{}^{-1} V_i V_i^*{}^{-1} X_i$ exists and is positive definite.

Condition 3'. Condition 3 holds when applied to $\Phi_i(\beta_0)$.

Condition 4'. Condition 4 holds.

It is easily seen that the covariance matrix defined in Condition 2' satisfies $\Sigma_3 = (\lim_{n \rightarrow \infty} N/n) \Sigma_1$. In the Appendix we show the following result.

THEOREM 3. *If Conditions 1'-4' hold, then as $n \rightarrow \infty$,*

$$-2 \log\{L_2(\beta_0)\} \rightarrow \chi_q^2$$

in distribution.

According to Qin & Lawless (1994, p. 305) the maximum empirical likelihood estimate $\widehat{\beta}$ satisfies the estimating equation (4) when V_i 's are known. Thus, if the V_i 's are properly estimated, both proposed empirical likelihoods lead to the same asymptotically efficient point estimator.

2.4. Inference for a linear combination of parameters

We now consider a linear combination $\theta = c^T \beta$ for any vector $c = (c_1, \dots, c_q)^T$. As an important special case, if only one of the components of c is not zero, this reduces to inference about a single parameter.

We discuss inference based on element-wise and subject-wise empirical likelihood approaches, respectively. Without loss of generality, we assume $c_1 \neq 0$. Denote $\gamma = (\theta, \beta_2, \dots, \beta_q)^T$ and

$$w_{ij}^T = (w_{ij1}, \dots, w_{ijq}) = \left(\frac{x_{ij1}}{c_1}, x_{ij2} - \frac{c_2}{c_1} x_{ij1}, \dots, x_{ijq} - \frac{c_q}{c_1} x_{ij1} \right).$$

Then model (1) can be rewritten as

$$y_{ij} = w_{ij}^T \gamma + \epsilon_{ij}. \quad (5)$$

Write $W_i^T = (w_{i1}, \dots, w_{im_i})$. Then model (5) can be written in matrix form as $Y_i = W_i \gamma + \epsilon_i$. The generalized estimating equation estimator of γ in model (5) is denoted by $\tilde{\gamma}_N = (\tilde{\theta}_N, \tilde{\beta}_{N2}, \dots, \tilde{\beta}_{Nq})^T = (\tilde{\theta}_N, \tilde{\beta}_{N(1)}^T)^T$. Therefore, $\tilde{\gamma}_N$ satisfies

$$\sum_{i=1}^n W_i^T \hat{V}_{in}^{-1} (Y_i - W_i \gamma) = 0. \quad (6)$$

Rewrite the $m_i \times q$ matrix W_i as $W_i = (R_{i1}, R_{i2}, \dots, R_{iq}) = (R_{i1}, R_{i(1)})$. Using the element-wise method, we construct an auxiliary vector as follows:

$$u_{ij}(\theta) = w_{ij1} \left\{ V_i^{*-1} (Y_i - R_{i(1)} \tilde{\beta}_{N(1)} - R_{i1} \theta) \right\}_j,$$

since $E\{u_{ij}(\theta_0)\} = 0$ when $\theta_0 = c^T \beta_0$. Denote

$$\hat{u}_{ij}(\theta) = w_{ij1} \left\{ \hat{V}_{in}^{-1} (Y_i - R_{i(1)} \tilde{\beta}_{N(1)} - R_{i1} \theta) \right\}_j.$$

When the \hat{V}_{in} are consistent estimators of the V_i , $V_i^* = V_i$ in $u_{ij}(\theta)$. The empirical log-likelihood ratio is

$$\tilde{l}_{N1}(\theta) = -2 \sup_{p_{ij}} \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \log(N p_{ij}) \mid p_{ij} \geq 0, \sum_{i=1}^n \sum_{j=1}^{m_i} p_{ij} = 1, \sum_{i=1}^n \sum_{j=1}^{m_i} p_{ij} \hat{u}_{ij}(\theta) = 0 \right\}.$$

Let $\tilde{\sigma}_{N1}^2(\theta)$ be a consistent estimator of the asymptotic variance of $N^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} u_{ij}(\theta)$ and $\tilde{\sigma}_{N2}^2(\theta) = N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \hat{u}_{ij}^2(\theta)$. Then the adjustment factor can be constructed as $\tilde{a}_{N1}(\theta) = \tilde{\sigma}_{N2}^2(\theta) / \tilde{\sigma}_{N1}^2(\theta)$, where the denominator $\tilde{\sigma}_{N1}^2(\theta)$ can be calculated in a fashion similar to that given in the remark for Theorem 5 below.

Similar to Theorem 2, we can show that $\tilde{l}_{N1}(\theta)$ with a suitable adjustment factor follows asymptotically a χ_1^2 distribution. The result is given below.

THEOREM 4. *When Conditions 1-4 hold, as $n \rightarrow \infty$,*

$$\tilde{l}_{N1}(\theta_0) = \left\{ N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} u_{ij}^2(\theta_0) \right\}^{-1} \left\{ N^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} u_{ij}(\theta_0) \right\}^2 + o_p(1)$$

and

$$\tilde{a}_{N1}(\theta_0) \tilde{l}_{N1}(\theta_0) = \tilde{\sigma}_{N1}^{-2}(\theta_0) \left\{ N^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} u_{ij}(\theta_0) \right\}^2 + o_p(1) \rightarrow \chi_1^2$$

in distribution.

For the subject-wise method, define $\hat{u}_i(\theta) = \sum_{j=1}^{m_i} \hat{u}_{ij}(\theta)$,

$$\tilde{l}_{n2}(\theta) = -2 \sup_{p_i} \left\{ \sum_{i=1}^n \log(n p_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{u}_i(\theta) = 0 \right\},$$

and $\tilde{a}_{n2}(\theta) = \tilde{v}_{n2}^2(\theta)/\tilde{v}_{n1}^2(\theta)$, where $\tilde{v}_{n1}^2(\theta)$ is a consistent estimator of the asymptotic variance of $n^{-1/2} \sum_{i=1}^n u_i(\theta)$ with $u_i(\theta) = \sum_{j=1}^{m_i} u_{ij}(\theta)$ and $\tilde{v}_{n2}^2(\theta) = n^{-1} \sum_{i=1}^n \tilde{u}_i^2(\theta)$. Then we have the following result that is parallel to Theorem 4.

THEOREM 5. *Under Conditions 1'-4', as $n \rightarrow \infty$,*

$$\tilde{a}_{n2}(\theta_0) \tilde{l}_{n2}(\theta_0) \rightarrow \chi_1^2$$

in distribution.

Remark. By Theorem 5, we can construct a confidence interval for θ_0 once $\tilde{v}_{n1}^2(\theta)$ is obtained. Denote by $(A)_1$ and $(A)_{(1,1)}$ the first column and the (1,1) element of a matrix A respectively. Let $\Gamma_n = n^{-1} \sum_{i=1}^n W_i^T V_i^{*-1} V_i V_i^{*-1} W_i$, $\Gamma_n^* = n^{-1} \sum_{i=1}^n W_i^T V_i^{*-1} W_i$, and $\hat{\Gamma}_n$ and $\hat{\Gamma}_n^*$ be their estimators respectively, such as $\hat{\Gamma}_n = n^{-1} \sum_{i=1}^n W_i^T \tilde{Z}_i(\theta) \tilde{Z}_i^T(\theta) W_i$ and $\hat{\Gamma}_n^* = n^{-1} \sum_{i=1}^n W_i^T \hat{V}_{in}^{-1} W_i$, where $\tilde{Z}_i(\theta)$ is $\tilde{Z}_i(\theta, \tilde{\beta}_{N(1)})$ after proper recentering. Then $\tilde{v}_{n1}^2(\theta)$ can be defined as $(\hat{\Gamma}_n^{*-1})_1^T \hat{\Gamma}_n (\hat{\Gamma}_n^{*-1})_1 \{(\hat{\Gamma}_n^*)_{(1,1)}\}^2$. A brief proof is given in the Appendix.

3. MOST POWERFUL TEST BASED ON THE SUBJECT-WISE EMPIRICAL LIKELIHOOD

In this section, we address local power, focusing on the subject-wise method. Let \mathcal{D} be a class of n symmetric positive definite $m_i \times m_i$ matrices $\{D_i\}_{i=1}^n$ such that for any $\{D_i\}_{i=1}^n \in \mathcal{D}$ both $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T D_i X_i$ and $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T D_i V_i D_i X_i$ exist. Thus $\{V_i^{*-1}\}_{i=1}^n \in \mathcal{D}$. We shall prove that the proposed subject-wise empirical likelihood based method is the locally most powerful test within the class \mathcal{D} . Notice that when $D_i = \sigma^2 I_{m_i}$ for all i , this reduces to You et al.'s method, while when $D_i = V_i^{*-1}$, the corresponding test is the proposed subject-wise empirical likelihood method.

In the remaining discussion in this section, for notational simplicity, we denote $D_i = V_i^{*-1}$. Suppose that we wish to test $H_0 : \beta = \beta_0$ versus a local alternative $H_a : \beta = \beta_0 + n^{-1/2} \delta$ for a bounded $\delta \in R^q$. Let $\zeta_{\delta i} = D_i X_i \delta = (\zeta_{\delta i1}, \zeta_{\delta i2}, \dots, \zeta_{\delta im_i})^T$. Then

$$\Phi_{ij}(\beta_0 + n^{-1/2} \delta) = \Phi_{ij}(\beta_0) + n^{-1/2} X_{ij} \{D_i X_i \delta\}_j = \Phi_{ij}(\beta_0) + n^{-1/2} X_{ij} \zeta_{\delta ij}.$$

Hence

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \Phi_i(\beta_0 + n^{-1/2} \delta) &= n^{-1/2} \sum_{i=1}^n \Phi_i(\beta_0) + n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} X_{ij} \zeta_{\delta ij} \\ &\rightarrow N(\Sigma_3^* \delta, \Sigma_3) \end{aligned}$$

in distribution, where Σ_3 is given in Condition 2' and $\Sigma_3^* = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T D_i X_i$. Since $\{D_i\}_{i=1}^n \in \mathcal{D}$, we have

$$n^{-1} \sum_{i=1}^n \Phi_i(\beta_0 + n^{-1/2} \delta) \Phi_i^T(\beta_0 + n^{-1/2} \delta) = \Sigma_3 + o_p(1).$$

Therefore, we have the following property for the corresponding empirical likelihood:

$$\begin{aligned} -2 \log L_2(\beta_0 + n^{-1/2} \delta) &\rightarrow (\mathcal{Z}_q + \Sigma_3^{-1/2} \Sigma_3^* \delta)^T \Sigma_3^{1/2} \Sigma_3^{-1} \Sigma_3^{1/2} (\mathcal{Z}_q + \Sigma_3^{-1/2} \Sigma_3^* \delta) \\ &= \chi_q^2(\delta^T \Sigma_3^* \Sigma_3^{-1} \Sigma_3^* \delta) = \chi_q^2(\Delta_D) \end{aligned}$$

in distribution, where $\Delta_D = \delta^T \Sigma_3^* \Sigma_3^{-1} \Sigma_3^* \delta$ is the non-centrality parameter and \mathcal{Z}_q is a vector of q independent standard normal random variables.

THEOREM 6. *Among the class of the estimators with all possible $\{D_i\}_{i=1}^n = \{V_i^{*-1}\}_{i=1}^n \in \mathcal{D}$, having $D_i = V_i^{-1}$ for all i gives the asymptotically locally most powerful test.*

A sketch proof of Theorem 6 is given in the Appendix.

4. A SIMULATION STUDY

In this section, we present a simulation study to investigate the finite sample properties of the two generalized empirical likelihood methods, comparing them with the normal approximation and You et al.'s (2006) block or subject-wise empirical likelihood method. As we noted at the end of §2.3, the point estimates of our two methods are the same as the generalized estimating equation method with the same covariance matrices, while the point estimate of You et al. is the same as the generalized estimating equation method with the working independence assumption. The comparisons of generalized least squares and ordinary least squares estimates in correlated data problems are well known, so here we confine attention to the coverage probabilities of the methods and the average widths of their confidence intervals.

Due to the nature of the within correlation structure, the issue narrows down to incorporating the within-subject covariance matrices in our procedures. We demonstrate the usefulness of the proposed methods with two possible options of handling the within-subject covariance matrices. They are: (a) the nonparametric sample covariance matrix obtained from the residuals of a generalized estimating equation fit using working independence and (b) the fitted covariance matrix under a correct specification of the correlation structure. We compared our proposed methods with You et al.'s subject-wise empirical likelihood method and the normal approximations using the generalized estimating equation method with working independence, and with (a) and (b) above.

We generated data from the same longitudinal model as in You et al., namely

$$y_{ij} = x_{ij}\beta + \epsilon_{ij}, \quad i = 1, \dots, n; j = 1, \dots, m_i,$$

where $x_{ij} \sim N(0, 1)$, $\beta = 3.5$ and $\epsilon_{ij} = be_i + (1 - b)e_{ij}$ with $e_i \sim N(0, 1)$, which is independent of $e_{ij} \sim N(0, c^2)$, and $0 < b < 1$.

Of course ϵ_{ij} has the exchangeable correlation structure with within-subject correlation coefficient $\rho = b^2 / \{b^2 + (1 - b)^2 c^2\}$. Hence considering various values of c reveals the efficiency of our methods compared to working independence methods such as You et al. and the normal approximation. In our simulation study, we considered $b = 0.5$ and three values of c : $c = 0.5, 1.0$ and 1.5 , representing a strong, moderate and weak within correlation structure with $\rho = 0.80, 0.50$, and 0.30 , respectively.

The numbers of subjects considered in the simulation were $n = 30, n = 50$ and $n = 100$. The number of repeated measurements for each subject is $m_i = 3$. For each combination of (n, ρ) , the simulation was repeated 1000 times. For each sample, the point estimate of β was computed using the generalized estimation equation $\sum_{i=1}^n \sum_{j=1}^{m_i} \hat{\Phi}_{ij}(\beta) = 0$ for both of the proposed element-wise and subject-wise methods.

The 95% coverage probability for β was computed for all methods. The normal approximation was constructed from the generalized estimating equation point estimate $\pm 1.96 \times$ the estimated standard error using the generalized estimating equation approach. The coverage probability for the empirical likelihood based methods was constructed using the asymptotic χ_1^2 distribution.

The package *geese* from *geepack* in R was used in computing the generalized estimating equation estimator and its standard error. The results of the simulation study are shown in Table 1. The results are in line with our expectations, as follows.

Table 1. Coverage probabilities and average confidence interval lengths for the exchangeable correlation case with varying amounts of correlation. The four methods compared are You et al.'s method, normal approximation (NA), element-wise empirical likelihood (EW) and subject-wise empirical likelihood (SW). The term “.res.ind” means using the residuals from a working independence fit to estimate the correlation matrix nonparametrically, “.exch” means estimation with the fitted correct exchangeable correlation structure, while “.exch.ar1” means estimation with the fitted misspecified autoregressive(1) correlation structure.

n	Method	$\rho = 0.80$		$\rho = 0.50$		$\rho = 0.30$	
		Coverage	Width	Coverage	Width	Coverage	Width
30	You et al.	92.5	0.231	93.4	0.290	93.8	0.367
	NA.ind	93.0	0.226	93.5	0.286	94.3	0.365
	NA.res.ind	90.5	0.117	91.4	0.225	92.3	0.325
	NA.exch	91.7	0.117	92.2	0.225	92.7	0.325
	EW.res.ind	93.2	0.128	92.9	0.239	93.5	0.343
	EW.exch	95.4	0.132	95.3	0.251	95.5	0.361
	EW.exch.ar1	95.5	0.146	95.3	0.267	95.6	0.371
	SW.res.ind	92.8	0.126	92.5	0.233	92.5	0.336
	SW.exch	93.6	0.128	93.7	0.245	94.1	0.354
SW.exch.ar1	93.8	0.139	94.1	0.258	94.3	0.367	
50	You et al.	93.4	0.179	92.5	0.227	94.2	0.289
	NA.ind	93.7	0.177	92.2	0.223	94.2	0.286
	NA.res.ind	92.3	0.094	92.5	0.180	92.9	0.259
	NA.exch	93.3	0.094	93.6	0.180	94.2	0.259
	EW.res.ind	94.0	0.100	93.8	0.189	94.4	0.275
	EW.exch	95.9	0.103	95.1	0.195	95.3	0.283
	EW.exch.ar1	95.5	0.112	96.1	0.206	95.4	0.288
	SW.res.ind	92.8	0.098	93.7	0.187	94.5	0.272
	SW.exch	94.9	0.101	94.9	0.193	95.3	0.280
SW.exch.ar1	94.9	0.110	95.5	0.203	95.2	0.286	
100	You et al.	93.4	0.128	94.9	0.161	94.8	0.205
	NA.ind	94.0	0.126	94.8	0.159	94.9	0.203
	NA.res.ind	94.2	0.067	94.0	0.129	94.0	0.187
	NA.exch	95.0	0.067	94.5	0.129	94.2	0.187
	EW.res.ind	95.7	0.071	94.9	0.137	95.5	0.197
	EW.exch	96.6	0.072	95.3	0.139	95.8	0.199
	EW.exch.ar1	95.8	0.079	96.1	0.146	95.0	0.206
	SW.res.ind	95.3	0.071	94.8	0.136	95.1	0.196
	SW.exch	96.2	0.072	95.4	0.138	95.6	0.199
SW.exch.ar1	95.2	0.079	95.6	0.146	95.0	0.208	

- (a) If the structure of the covariance matrix is correctly specified and the maximum likelihood method is used to estimate it, our methods produced coverage probabilities roughly at the nominal level along with significantly shorter average lengths than the methods with working independence, especially when the correlation is high, $\rho = 0.80$. When the correlation is modest, $\rho = 0.50$, the confidence interval lengths are still 15% – 20% shorter.
- (b) Both the working independence methods and the normal approximations with the exchangeability assumption have some under-coverage for $n = 30$ and $n = 50$.
- (c) When the covariance matrix is estimated nonparametrically from the residuals, our empirical likelihood methods have coverage similar to the working independence methods, but they maintain their advantage in average confidence interval length.
- (d) While the two generalized empirical likelihood methods generally performed similarly, the element-wise empirical likelihood method tends to give slightly better coverage probabilities for small to moderate sample sizes.

To examine numerically the effect of using a misspecified correlation structure in the empirical methods, we have also included in Table 1 the results when an autoregressive correlation structure was assumed while in fact the true correlation structure is exchangeable. The misspecification appears to have little impact. Simulations were also carried out when using misspecified exchangeable structure to fit data simulated from an autoregressive structure with similar results. In addition, similar results were obtained in the case of $m_i = 4$. Another special misspecified case is of course the You et al. method.

Additional empirical analyses using various cluster sizes were conducted and gave similar results.

5. APPLICATION TO THE FRAMINGHAM HEART STUDY

A well-known ongoing longitudinal study is the Framingham Heart Study. This study has produced fifty years of data collected from residents of Framingham and identified major risk factors associated with heart disease, stroke and other diseases. Further details on the background of the study can be found online at <http://www.framingham.com/heart/backgrnd.htm>. The resulting participant clinic data set includes participant identification number, sex, age, body mass index and systolic blood pressure and is organized in a longitudinal form.

Suppose that we are interested in how systolic blood pressure relates to either age or body mass index. The response variable y_{ij} considered here is systolic blood pressure. The covariate x_{ij} considered is age or body mass index of the participants. First we consider the following simple linear regression model:

$$y_{ij} = \beta_1 + \beta_2 x_{ij} + \epsilon_{ij},$$

where β_1 and β_2 describe the baseline systolic blood pressure and the effect of age or body mass index, respectively.

We computed nonparametrically estimated correlation matrices for systolic blood pressure after controlling for age and body mass index respectively, using 3182 participants collected during three examination periods, approximately 6 years apart from roughly 1956 to 1968. They suggest moderate to strong within-subject correlations and that the correlations might be approximated by an exchangeable structure with a constant variance over time. For illustration, we extracted a random subsample of $n = 30$ patients and multiplied both the response variable and the covariate by a factor of 0.01 to view the confidence region better.

We computed four likelihood ratios: You et al.'s method, the normal approximation, element-wise and subject-wise with exchangeable correlation structure. The confidence regions were computed according the theoretical results in §2. Let $\chi_q^2(\alpha)$ be the $1 - \alpha$ quantile of χ_q^2 for $0 < \alpha < 1$. Then Theorem 2 implies that for the element-wise method, an approximate $100(1 - \alpha)\%$ confidence region for $\beta = (\beta_1, \beta_2)^T$ is

$$R_\alpha^{EW} = \{\beta : -2a_N(\beta) \log\{L_1(\beta)\} \leq \chi_q^2(\alpha)\}.$$

For the subject-wise method, an approximate $100(1 - \alpha)\%$ confidence region for β is

$$R_\alpha^{SW} = \{\beta : -2 \log\{L_2(\beta)\} \leq \chi_q^2(\alpha)\}.$$

The 95% confidence regions are reported in Fig. 1 for body mass index. As expected, this figure shows that accounting for the correlation leads to significantly smaller confidence regions than using working independence. You et al. produces much wider confidence region, while the other three methods are generally comparable in this case.

However, the simulation results in §4 suggest that the normal approximation coverage with correct exchangeable correlation structure tends to be lower than the target value for small to moderate sample sizes. We considered various sample sizes other than $n = 30$, and the resulting comparisons were similar. In addition, similar results were obtained for age as the covariate, but we omit the details.

Next, we illustrate the proposed methods for making inferences for a single parameter in a multiple regression setting described in §2.4. Here we study the multiple linear regression of systolic blood pressure on both age and body mass index simultaneously. Recall that inverting a matrix is required in this case of linear combination of parameters. Since for a small to moderate sample sizes this step might pose problems for You et al. and to a much lesser extent for the subject-wise method, we used a subsample of size $n = 100$ for more reliable comparisons. The 95% confidence interval limits of the regression parameters for the four estimation methods are reported in Table 2.

Table 2. *The 95% confidence limits for regression parameters using four methods of estimation for the Framingham Heart Study. The four methods compared are You et al.'s method, normal approximation, element-wise empirical likelihood and subject-wise empirical likelihood, all fitted under an exchangeable correlation structure. Here BMI is body mass index.*

Method	Covariates					
	Constant	AGE	BMI	AGE	BMI	BMI
You et al.	-9.40	117.13	0.15	2.02	-1.12	2.88
Normal Approximation	35.78	84.85	0.71	1.30	-0.08	1.68
Element-wise	40.89	79.07	0.75	1.26	0.05	1.45
Subject-wise	44.95	78.34	0.79	1.27	0.23	1.44

Table 2 shows that You et al. gives much longer confidence intervals than the other three methods. The element-wise and subject-wise methods give slightly shorter confidence intervals than the normal approximation. The increased length of the You et al. method is partly caused by estimating the first component κ_1 of the τ in their method with the residuals. If the estimated exchangeable correlation matrix were used when estimating κ_1 , the confidence intervals for age and body mass index become $(0.75, 1.60)$ and $(-0.11, 2.43)$, respectively. At test level 0.05, the element-wise and subject-wise methods show statistically significant effects of age and body mass index, while the normal approximation and the method of You et al. show the significance of age only.

It is worth pointing out that computation of empirical likelihood methods is generally non-trivial and can be a major obstacle especially when making simultaneous inferences for multiple regression coefficients. This potential problem is much reduced when making inferences on a linear combination of the parameters.

Finally, we wish to emphasize once more that we are only assuming that the covariance estimators \widehat{V}_{in} converge uniformly to some covariance matrices V_i^* , which may or may not be the correct covariance matrices V_i . It is generally satisfied if a purported common model for all V_i is assumed and it depends on a fixed number of unknown parameters, in which case the residuals from a working independence fit may be used to estimate these parameters.

6. DISCUSSION

To achieve full efficiency, our methods require the availability of consistently estimated within-subject correlation matrices for all subjects, although the methods are consistent and give correct asymptotic inference as long as the estimated correlation matrices have asymptotic limits. In some cases, such as in our simulation study, the sample residual covariance matrix is useful for this purpose. However, in many applications, more sophisticated covariance estimation procedures are desired. For example, Ye & Pan (2006) proposed to model covariance structures for generalized estimating equations for longitudinal data. How well our empirical likelihood methods combined with such covariance estimation procedures behave in various useful settings is an interesting topic for future research. In addition, extensions of our methods to more complicated regression models such as partially linear models will be considered elsewhere.

ACKNOWLEDGMENTS

We thank the Editor, the Associate Editor and two referees for their helpful comments and suggestions that have led to significant improvements of this paper. Carroll's research was supported by a grant from the National Cancer Institute and by a research award made by King Abdullah University of Science and Technology.

APPENDIX

Proofs and complements

Proof of Theorem 1. We only outline the main steps of the proof. We begin the proof by showing that using \widehat{G}_{in} or V_i^{*-1} leads to the same limit distribution. More generally, this result applies to other settings throughout the paper. By standard calculations, this is true once we show that

$$A_n = n^{-1/2} \sum_{i=1}^n \widehat{T}_{in} = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} (\widehat{g}_{ijk} - g_{ijk}) X_{ij} \epsilon_{ik} = o_p(1), \quad (\text{A1})$$

where $\widehat{T}_{in} = \sum_{j=1}^{m_i} \{\widehat{\Phi}_{ij}(\beta_0) - \Phi_{ij}(\beta_0)\}$. For given $i \neq i'$, partition $\widehat{T}_{in} = \widehat{B}_{ii'} + \widehat{C}_{ii'}$ with $\widehat{C}_{ii'} = \sum_{j=1}^{m_i} \sum_{k=1}^{m_{i'}} (\widehat{g}_{-(i,i')jk} - g_{ijk}) X_{ij} \epsilon_{ik} = o_p(1)$ uniformly by Condition 4. Then we have $E(A_n) = o(1)$ since $E(\widehat{C}_{ii'}) = 0$ and $E(\widehat{B}_{ii'}) = O(n^{-1})$. Moreover, $E(\widehat{C}_{ii'} \widehat{C}_{ii'}^T) = \sum_{j_1=1}^{m_i} \sum_{k_1=1}^{m_{i'}} \sum_{j_2=1}^{m_{i'}} \sum_{k_2=1}^{m_{i'}} E\{(\widehat{g}_{-(i,i')j_1k_1} - g_{ij_1k_1})(\widehat{g}_{-(i,i')j_2k_2} - g_{i'j_2k_2})\} X_{ij_1} X_{i'j_2}^T E(\epsilon_{ik_1} \epsilon_{i'k_2}) = 0$. By the Chebyshev inequality, for (A1) to be true it suffices to show that $\text{var}(A_n) = o(1)$. Observe that

$E(\widehat{T}_{in}\widehat{T}_{in}^T) = o(1)$. Therefore,

$$\begin{aligned}\text{var}(A_n) &= n^{-1} \sum_{i \neq i'} E(\widehat{T}_{in}\widehat{T}_{i'n}^T) + o(1) \\ &= n^{-1} \sum_{i \neq i'} E(\widehat{B}_{ii'}\widehat{B}_{i'i}^T + \widehat{B}_{ii'}\widehat{C}_{i'i}^T + \widehat{C}_{ii'}\widehat{B}_{i'i}^T + \widehat{C}_{ii'}\widehat{C}_{i'i}^T) + o(1) \\ &\equiv \Delta_{1n} + \Delta_{2n} + \Delta_{3n} + \Delta_{4n} + o(1) = o(1)\end{aligned}$$

since $\Delta_{4n} = 0$ and all of Δ_{1n} , Δ_{2n} and Δ_{3n} are of order $o(1)$ or smaller. Approximation (A1) has thus been shown.

It is easily seen by the Lagrange multipliers that Condition 1 implies that the unique weights $(\widehat{p}_{i1}, \dots, \widehat{p}_{im_i})$ that define $L_1(\beta_0)$ satisfy

$$\widehat{p}_{ij} = \frac{1}{N\{1 + \lambda^T \widehat{\Phi}_{ij}(\beta_0)\}} \text{ for all } i, j,$$

where $\lambda^T = (\lambda_1, \dots, \lambda_q)$ solves

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \frac{\widehat{\Phi}_{ij}(\beta_0)}{1 + \lambda^T \widehat{\Phi}_{ij}(\beta_0)} = 0.$$

With Conditions 2 and 3 and approximation (A1), one can show that

$$\lambda = \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) \Phi_{ij}^T(\beta_0) \right\}^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) + o_p(N^{-1/2})$$

and $\sum_{i=1}^n \sum_{j=1}^{m_i} \lambda^T \Phi_{ij}(\beta_0) = \sum_{i=1}^n \sum_{j=1}^{m_i} \{\lambda^T \Phi_{ij}(\beta_0)\}^2 + o_p(1)$. Then we have

$$\begin{aligned}-2 \log\{L_1(\beta_0)\} &= 2 \sum_{i=1}^n \sum_{j=1}^{m_i} \left[\lambda^T \Phi_{ij}(\beta_0) - \frac{1}{2} \{\lambda^T \Phi_{ij}(\beta_0)\}^2 \right] + o_p(1) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \lambda^T \Phi_{ij}(\beta_0) \Phi_{ij}^T(\beta_0) \lambda + o_p(1) \\ &= \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) \right\}^T \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) \Phi_{ij}^T(\beta_0) \right\}^{-1} \\ &\quad \times \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) + o_p(1).\end{aligned}\tag{A2}$$

Now, as $n \rightarrow \infty$ and by Condition 2,

$$\begin{aligned}N^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \Phi_{ij}(\beta_0) &= N^{-1/2} \sum_{i=1}^n X_i^T Z_i(\beta_0) = N^{-1/2} \sum_{i=1}^n X_i^T V_i^{*-1} (Y_i - X_i \beta_0) \\ &\rightarrow \mathbf{N}(0, \Sigma_1)\end{aligned}$$

in distribution. Therefore, by Condition 2, as $n \rightarrow \infty$, $-2 \log\{L_1(\beta_0)\} \rightarrow \mathcal{Z}_q^T \Sigma_1^{1/2} \Sigma_2^{-1} \Sigma_1^{1/2} \mathcal{Z}_q = \sum_{k=1}^q c_k \chi_{1k}^2$ in distribution, where \mathcal{Z}_q is a vector of q independent standard normal random variables, completing the proof.

Proof of Theorem 3. The proof is similar to that of Theorem 1. First, by the Lagrange multipliers, Condition 1' implies that there is a unique set of weights:

$$\widehat{p}_i = \frac{1}{n\{1 + \tau^T \widehat{\Phi}_i(\beta_0)\}}$$

in $L_2(\beta_0)$ for all i , where $\tau^T = (\tau_1, \dots, \tau_q)$ solves

$$\sum_{i=1}^n \frac{\widehat{\Phi}_i(\beta_0)}{1 + \tau^T \widehat{\Phi}_i(\beta_0)} = 0.$$

Then by Conditions 2'-4', we also have $\tau = \left\{ \sum_{i=1}^n \Phi_i(\beta_0) \Phi_i^T(\beta_0) \right\}^{-1} \sum_{i=1}^n \Phi_i(\beta_0) + o_p(n^{-1/2})$ and $\sum_{i=1}^n \tau^T \Phi_i(\beta_0) = \sum_{i=1}^n \left\{ \tau^T \Phi_i(\beta_0) \right\}^2 + o_p(1)$. Thus,

$$-2 \log \{L_2(\beta_0)\} = \left\{ \sum_{i=1}^n \Phi_i(\beta_0) \right\}^T \left\{ \sum_{i=1}^n \Phi_i(\beta_0) \Phi_i^T(\beta_0) \right\}^{-1} \sum_{i=1}^n \Phi_i(\beta_0) + o_p(1).$$

By Condition 2', as $n \rightarrow \infty$, $n^{-1/2} \sum_{i=1}^n \Phi_i(\beta_0) \rightarrow N(0, \Sigma_3)$ in distribution, where $\Sigma_3 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \text{cov}\{\Phi_i(\beta_0)\} = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T V_i^{*-1} V_i V_i^{*-1} X_i$ was defined in Condition 2'. On the other hand, $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \Phi_i(\beta_0) \Phi_i^T(\beta_0) = \Sigma_3$, completing the proof.

Formula for $\widehat{v}_{n1}^2(\theta_0)$ in Theorem 5. Similar to the proof of Theorem 3, one can prove that

$$\tilde{l}_{n2}(\theta) = \left\{ n^{-1} \sum_{i=1}^n u_i^2(\theta) \right\}^{-1} \left\{ n^{-1/2} \sum_{i=1}^n u_i(\theta) \right\}^2 + o_p(1),$$

where $u_i(\theta) = R_{i1}^T V_i^{*-1} [Y_i - R_{i(1)} \tilde{\beta}_{N(1)} - R_{i1} \theta]$. Here $\tilde{\gamma}_N$ satisfies (6). Hence,

$$\tilde{\gamma}_N - \gamma = n^{-1} \Gamma_n^{*-1} \sum_{i=1}^n W_i^T V_i^{*-1} \epsilon_i + o_p(n^{-1/2}) \quad \text{and} \quad \tilde{\theta}_N - \theta = n^{-1} (\Gamma_n^{*-1})_1^T \sum_{i=1}^n W_i^T V_i^{*-1} \epsilon_i + o_p(n^{-1/2}).$$

Now rewrite $n^{-1/2} \sum_{i=1}^n u_i(\theta)$ as

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n u_i(\theta) &= n^{-1/2} \sum_{i=1}^n R_{i1}^T V_i^{*-1} \left\{ (Y_i - W_i \gamma) + W_i (\gamma - \tilde{\gamma}_N) + R_{i1} (\tilde{\theta}_N - \theta) \right\} \\ &\equiv \Psi_{1n} + \Psi_{2n} + \Psi_{3n}. \end{aligned}$$

Then the variance of $n^{-1/2} \sum_{i=1}^n u_i(\theta_0)$ is the sum of the following six terms: the three variances terms

$$\begin{aligned} \text{var}(\Psi_{1n}) &= n^{-1} \sum_{i=1}^n R_{i1}^T V_i^{*-1} V_i V_i^{*-1} R_{i1} = (\Gamma_n)_{(1,1)}, \\ \text{var}(\Psi_{2n}) &= (\Gamma_n^*)_1^T \Gamma_n^{*-1} \Gamma_n \Gamma_n^{*-1} (\Gamma_n^*)_1 + o(1) = (\Gamma_n)_{(1,1)} + o(1), \\ \text{var}(\Psi_{3n}) &= (\Gamma_n^{*-1})_1^T \Gamma_n (\Gamma_n^{*-1})_1 \{(\Gamma_n^*)_{(1,1)}\}^2 + o(1), \end{aligned}$$

and the three covariance terms

$$\begin{aligned} 2\text{cov}(\Psi_{1n}, \Psi_{2n}) &= -2(\Gamma_n)_1^T \Gamma_n^{*-1} (\Gamma_n^*)_1 + o(1) = -2(\Gamma_n)_{(1,1)} + o(1), \\ 2\text{cov}(\Psi_{1n}, \Psi_{3n}) &= 2(\Gamma_n)_1^T (\Gamma_n^{*-1})_1 (\Gamma_n^*)_1 + o(1) = 2(\Gamma_n)_{(1,1)} + o(1), \\ 2\text{cov}(\Psi_{2n}, \Psi_{3n}) &= -2(\Gamma_n^*)_1^T \Gamma_n^{*-1} \Gamma_n \Gamma_n^{*-1} (\Gamma_n^*)_1 + o(1) = -2(\Gamma_n)_{(1,1)} + o(1). \end{aligned}$$

Thus, $\text{var}\{n^{-1/2} \sum_{i=1}^n u_i(\theta_0)\} = (\Gamma_n^{*-1})_1^T \Gamma_n (\Gamma_n^{*-1})_1 \{(\Gamma_n^*)_{(1,1)}\}^2 + o(1)$ which reduces to $(\Gamma_n^{-1})_{(1,1)} \{(\Gamma_n)_{(1,1)}\}^2 + o(1)$ when all $V_i^* = V_i$.

Proof of Theorem 6. For convenience of reference in proving the efficiency of the subject-wise method, we include the following generalized Cauchy-Schwartz inequality, see §B.10.2.2 of Bickel & Doksum, 2001.

LEMMA A1. If $\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$ is a symmetric semi-positive definite matrix, then $C_{11} \geq C_{12}C_{22}^{-1}C_{21}$, where “ \geq ” means the difference $C_{11} - C_{12}C_{22}^{-1}C_{21}$ is semi-positive definite.

Observe that

$$\begin{aligned} & \begin{pmatrix} \sum_{i=1}^n X_i^T V_i^{-1} X_i & \sum_{i=1}^n X_i^T D_i X_i \\ \sum_{i=1}^n X_i^T D_i X_i & \sum_{i=1}^n X_i^T D_i V_i D_i X_i \end{pmatrix} \\ &= \sum_{i=1}^n X_i^T (D_i, D_i V_i D_i)^T D_i^{-1} V_i^{-1} D_i^{-1} (D_i, D_i V_i D_i) X_i \geq 0, \end{aligned}$$

so that the matrix is semi-positive definite. Then by the generalized Cauchy-Schwartz inequality in Lemma A1, $\sum_{i=1}^n X_i^T V_i^{-1} X_i \geq (\sum_{i=1}^n X_i^T D_i X_i) (\sum_{i=1}^n X_i^T D_i V_i D_i X_i)^{-1} (\sum_{i=1}^n X_i^T D_i X_i)$. Dividing both sides by n and letting $n \rightarrow \infty$, we obtain $\Sigma_3^{(0)} \geq \Sigma_3^* \Sigma_3^{-1} \Sigma_3^*$, where $\Sigma_3^{(0)} = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n X_i^T V_i^{-1} X_i$. When $D_i = V_i^{-1}$ for all i , the equation holds. That is, Δ_D is maximized when $D_i = V_i^{-1}$ in which case the corresponding test is asymptotically most powerful, completing the proof.

REFERENCES

- BICKEL, P. J. & DOKSUM, K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. **1**, 2nd ed. New York: Prentice Hall.
- LI, G. & WANG, Q. H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13**, 51-68.
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **74**, 237-249.
- OWEN, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall/CRC.
- QIN, J. & LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.
- RAO, J. N. K. & SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Am. Statist. Assoc.* **76**, 221-230.
- XUE, L. G. & ZHU, L. X. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika* **94**, 921-937.
- YE, H. & PAN, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data, *Biometrika* **93**, 927-941.
- YOU, J., CHEN, G. & ZHOU, Y. (2006). Block empirical likelihood for longitudinal partially linear regression models. *Can. J. Statist.* **34**, 79-96.
- ZHAO, Y. (2005). Regression analysis for long-term survival rate via empirical likelihood. *J. Nonparam. Statist.* **17**, 995-1007.
- ZOU, F., FINE, J. P. & YANDELL, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika* **89**, 61-75.

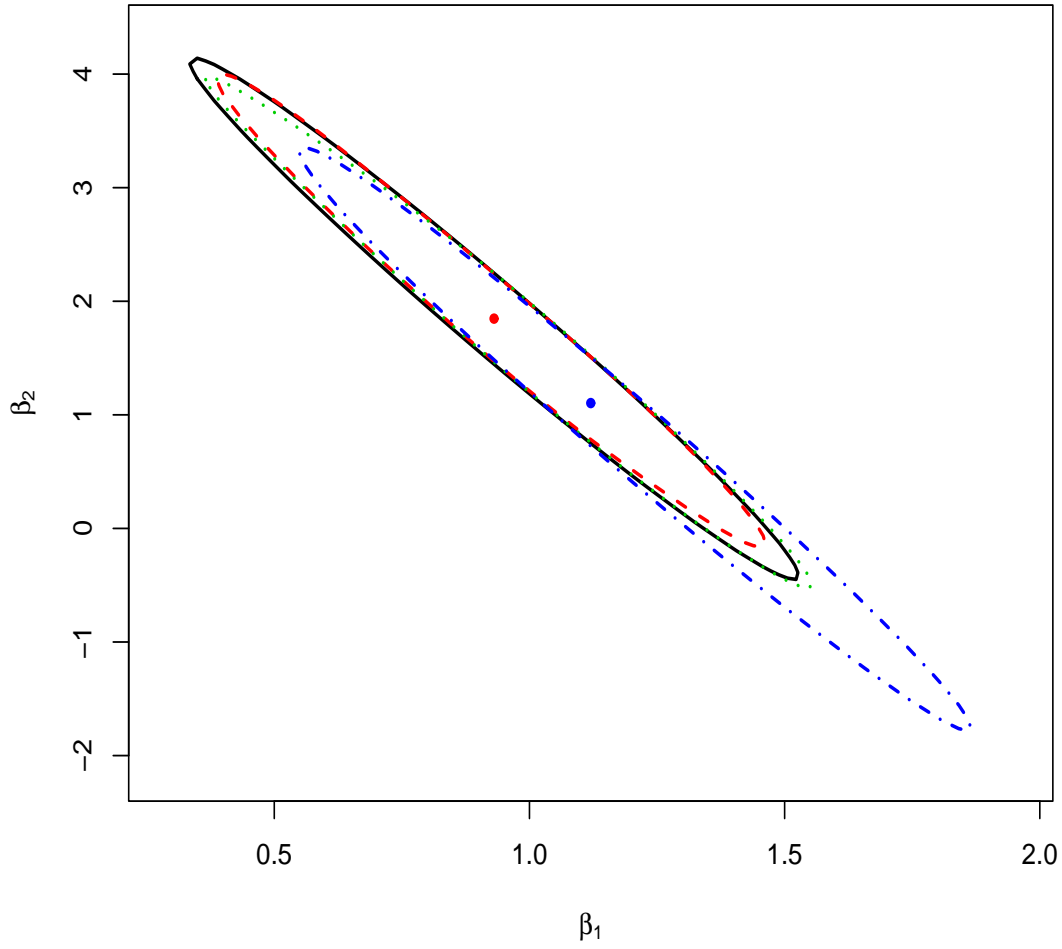


Fig. 1. The 95% confidence regions for (β_1, β_2) , where β_1 is the intercept and β_2 is the coefficient for body mass index. The four methods compared are You et al.'s method (dot-dashed line), normal approximation (solid line), element-wise empirical likelihood (dashed line) and subject-wise empirical likelihood (dotted line) with fitted correct exchangeable correlation structure (NA.exch, EW.exch and SW.exch, respectively).