

On weighted hybrid pairwise likelihood estimation for longitudinal data with LEAR correlation

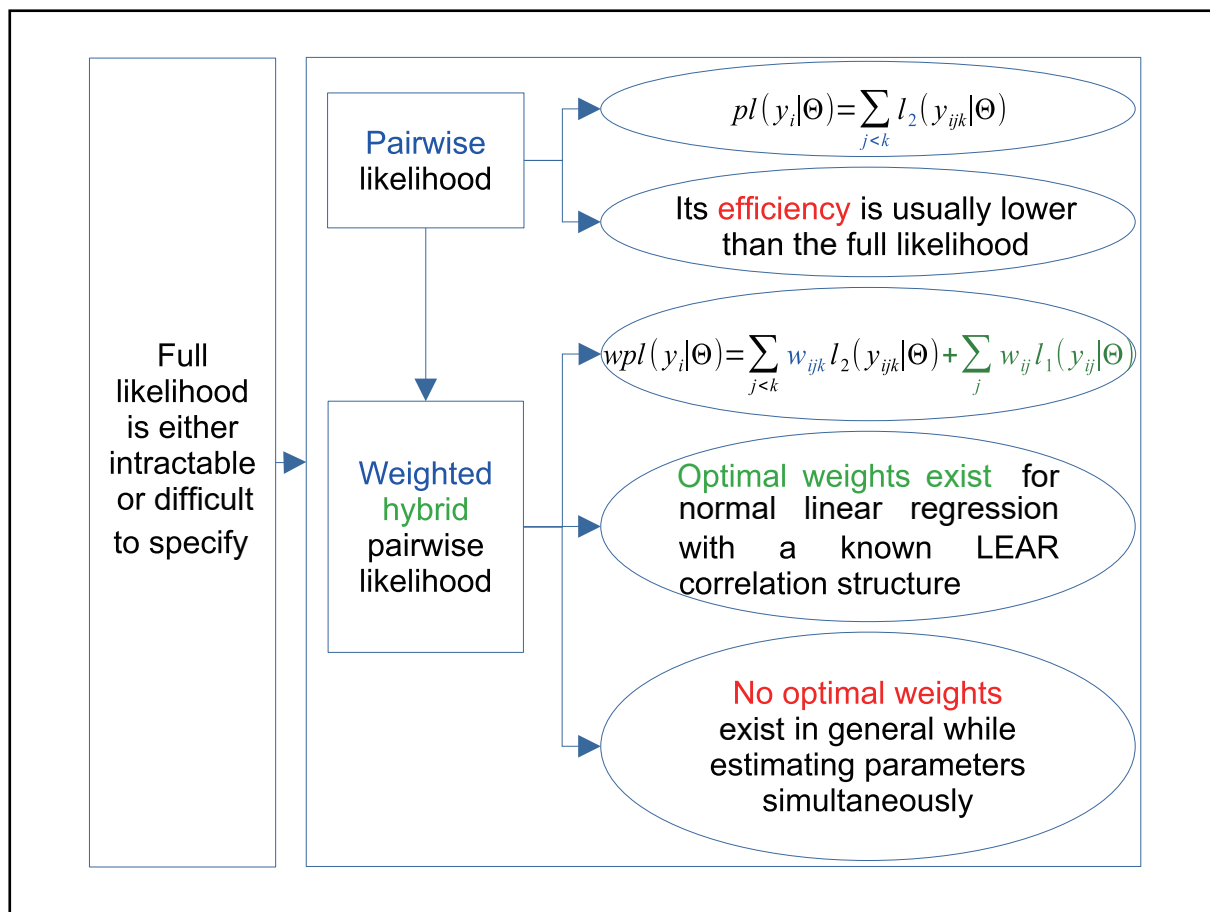
Feiyue Xie, Xiao Lin, and Weiping Zhang ✉

School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Weiping Zhang, E-mail: zwp@ustc.edu.cn

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Graphical abstract



Research framework.

Public summary

- We study the optimal weights for the weighted hybrid pairwise likelihood approach.
- We obtain some analytic results under the multivariate normal distribution with a linear exponent autoregressive (LEAR) correlation structure.
- We theoretically address the existence of optimal weights under several cases.

On weighted hybrid pairwise likelihood estimation for longitudinal data with LEAR correlation

Feiyue Xie, Xiao Lin, and Weiping Zhang 

School of Management, University of Science and Technology of China, Hefei 230026, China

 Correspondence: Weiping Zhang, E-mail: zwp@ustc.edu.cn

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Cite This: *JUSTC*, 2026, 56(2): 0201 (12pp)



Read Online

Abstract: Pairwise likelihood estimation is widely used to address the intractable full likelihood functions or the difficulty in specifying and verifying the correct high-dimensional joint distribution of data. Weighting has been proposed for increasing efficiency but the choice of weights is not straightforward in most applications. In this paper, we discuss the weight selection for hybrid pairwise likelihood estimation in longitudinal data. Some analytic results are obtained under a multivariate normal distribution with a linear exponent autoregressive (LEAR) correlation structure. We demonstrate the performance of different weighting schemes via simulations and analysis of CD4 data.

Keywords: optimal weights; repeated measurements; autoregressive correlation

CLC number: O212.1

Document code: A

2020 Mathematics Subject Classification: 62F12

1 Introduction

Longitudinal studies often involve repeated measurements of multiple outcomes over time on a group of subjects. The correlated nature and types of longitudinal data often make it difficult to evaluate, as well as to verify, the full likelihood function. For example, modeling longitudinal count data via latent variable models commonly involves high-dimensional integrals in likelihood, which is costly to evaluate or even intractable^[1,2]. To address this challenge, one can maximize a sum of low-dimensional marginal log-likelihoods, known as a composite likelihood, instead of the full likelihood^[3]. Typically, when employing bivariate marginals, we use the term “pairwise likelihood” for the composite likelihood. Moreover, composite likelihood methods are not restricted to specific data types and thus can be applied more broadly^[4].

Pairwise likelihood methods are widely applied in longitudinal data analysis. Diagnostic tools, such as correlation plots of various types and the lorelogram^[5], have been developed to check pairwise association structures, yet few methods are currently available for checking higher-order structures exist. For example, Kuk et al.^[6] discussed the computational advantages of pairwise likelihood methods relative to competing approaches for longitudinal binary data. Tang et al.^[1] and Zhang et al.^[2] utilized pairwise likelihood methods to ease the computational burden in analyzing generic discrete longitudinal data where high-dimensional integrals are involved in the full log-likelihood. Since the expectation of each likelihood component has a global maximum at the true parameters, the maximum pairwise likelihood estimator is guaranteed to be asymptotically normal under some conditions^[7]. The primary advantage of this approach is its computational efficiency since it avoids computing the full

likelihood. If the calculation resources are limited, Mazo et al.^[8] attempted to decrease the calculation cost by introducing a randomized Bernoulli weighting method. Owing to the neglect of correlations, the efficiency of composite likelihood estimation is usually lower than that of traditional full likelihood estimation.

To alleviate the inefficiencies caused by pairwise likelihood, various weighting techniques have been proposed to increase efficiency^[9,10]. There are two types of weighting schemes in the literature. The first type is component weights which are used to adjust the contribution of each component of pairwise log-likelihood or score functions. In practice, studies that do not focus on weights (e.g., Ref. [11]) tend to set most of the component weights to one, while setting weights of components with negligible information contribution to zero. This method may provide smaller asymptotic variances for estimators of the parameters of main interest, but may also result in slightly larger asymptotic variances for the estimators of other parameters, which suggests that these ad hoc weights may not be optimal. In fact, it might be better to use different weights for different parameters, that is, using weighted estimation equation approaches rather than maximizing weighted pairwise likelihood would yield estimators with smaller asymptotic variances. Kuk et al.^[12] first noted this issue and proposed a hybrid pairwise likelihood approach that combines the pairwise likelihood score functions with the optimal linear combinations of the marginal score functions. Lindsay et al.^[13] suggested converting the composite likelihood into an estimation equation framework to obtain optimal weights. Taking a step further, Vasdekis et al.^[14], inspired by Taylor expansion, viewed the pairwise likelihood estimator as a weighted estimator of estimates obtained from

separate maximizations of pairwise log-likelihoods, each summed across subjects. The second type is the cluster weight which is used to balance the information derived from different lengths of cluster or longitudinal observations. Joe and Lee^[10] compared the asymptotic relative efficiencies of different cluster-weighted component likelihood estimators, and reported that the best choice of weights as a function of cluster size depends on the observation count of each individuals, while the optimal weights may not the same for each parameter. Papageorgiou et al.^[15] also discussed various weight selection options and reached the same conclusion.

In this paper, we consider the optimal weights for the weighted hybrid pairwise likelihood approach. Some analytic results are obtained via the multivariate normal distribution with the linear exponent autoregressive (LEAR) correlation structure, which is a two-parameter generalization of the continuous-time AR(1) model^[16]. Special cases of the LEAR family include the AR(1), compound symmetry, and first-order moving average correlation structures, which are commonly used in longitudinal data analysis. The asymptotic normality of the estimator is also investigated in the general composite likelihood framework.

The remainder of this paper is organized as follows. Section 2 presents the weighted hybrid pairwise likelihood approach. The explicit estimators and their covariance matrices are discussed under a multivariate normal model with a LEAR correlation structure. Section 3 conducts numerical simulations to evaluate the performance of different weighting schemes. Section 4 demonstrates the advantages of the proposed method by applying it to the CD4 dataset and conducting comparisons with several other algorithms. Finally, Section 5 summarizes the entire work and draws conclusions.

2 Methodology

For $i = 1, \dots, n$, let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})^T$ be the m_i longitudinal observations for the i th subject measured at time $\mathbf{t}_i = (t_{i1}, \dots, t_{im_i})^T$, where m_i takes value from 1 to m with $m \geq 1$. That is, the longitudinal data could be unbalanced. We assume that the bivariate density $f_2(y_{ij}, y_{ik}; \boldsymbol{\theta})$ of Y_{ij} and Y_{ik} for $j \neq k$ and univariate marginal density $f(y_{ij}; \boldsymbol{\theta})$ of y_{ij} can be specified parametrically in terms of a vector of parameters $\boldsymbol{\theta}$. The following weighted hybrid pairwise log-likelihood function is commonly used for parameter estimation when the full likelihood is intractable or difficult to specify,

$$cl(\boldsymbol{\theta}; \mathbf{w}, \mathbf{y}) = \sum_{i=1}^n \left[\sum_{j < k} w_{ijk} \ln f_2(y_{ij}, y_{ik}; \boldsymbol{\theta}) + \sum_j w_{ijj} \ln f(y_{ij}; \boldsymbol{\theta}) \right] = \sum_{i=1}^n \left[\frac{1}{2} \sum_{j \neq k} w_{ijk} \ln f_2(y_{ij}, y_{ik}; \boldsymbol{\theta}) + \sum_j w_{ijj} \ln f(y_{ij}; \boldsymbol{\theta}) \right], \quad (1)$$

where $\mathbf{w} = \{w_{ijk}\}$ is a set of weights chosen by users if the optimal weights are difficult to specify, and $\mathbf{y}$ denotes all observations. The pairwise weights $w_{ijk} (j \neq k)$ are typically non-negative and symmetric $w_{ijk} = w_{ikj}$, whereas the marginal weights w_{ijj} are allowed to be negative to include the pairwise conditional likelihood^[17] in our framework. Note that for

the composite likelihood approach, the parameter vector should be identifiable from the set of bivariate margins.

It has been proven that the maximum composite likelihood estimator $\hat{\boldsymbol{\theta}}_{cl}$ of Eq. (1), given that the weights $\mathbf{w}$ are fixed, asymptotically follows normality^[17,18]:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{cl} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{G}_w^{-1}(\boldsymbol{\theta})), \quad (2)$$

where $\mathbf{G}_w(\boldsymbol{\theta}) = \mathbf{H}_w^T(\boldsymbol{\theta}) \mathbf{V}_w^{-1}(\boldsymbol{\theta}) \mathbf{H}_w(\boldsymbol{\theta})$ with

$$\mathbf{H}_w(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} E\{-\nabla_{\boldsymbol{\theta}}^2 cl(\boldsymbol{\theta}; \mathbf{w}, \mathbf{y})\}$$

and

$$\mathbf{V}_w(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}\{\nabla_{\boldsymbol{\theta}} cl(\boldsymbol{\theta}; \mathbf{w}, \mathbf{y})\}.$$

Here, $\nabla_{\boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\theta}}^2$ denote the first and second partial derivatives with respect to $\boldsymbol{\theta}$, respectively. In general, the identity $\mathbf{H}_w(\boldsymbol{\theta}) = \mathbf{V}_w(\boldsymbol{\theta})$ does not hold in the case of composite likelihoods. The assumed independence among the likelihood components forming the composite likelihood is usually not valid when the full likelihood is considered for longitudinal data.

To discuss the optimal weighting schemes, we study the weighted hybrid pairwise likelihood estimation for the multivariate normal distribution with the LEAR correlation structure, which includes several commonly used correlation models in longitudinal data analysis. We show that a hybrid pairwise likelihood estimation can be as efficient as the full likelihood estimation in some situations, whereas the asymptotic relative efficiency (ARE) can be poor in some other situations; we indicate situations in which this occurs. From now, let $\mathbf{Y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \sigma^2 \mathbf{R}_i(\rho, \delta))$ with $m_i \times p$ covariate matrix $\mathbf{X}_i = (X_{i1}, \dots, X_{im_i})^T$ and LEAR correlation structure^[16]:

$$\mathbf{R}_{i,jk}(\rho, \delta) = \begin{cases} \rho^{d_{\min} + \delta[(d_{ijk} - d_{\min})/(d_{\max} - d_{\min})]}, & j \neq k; \\ 1, & j = k, \end{cases} \quad (3)$$

where $\boldsymbol{\beta}$ is the regression coefficient and $d_{ijk} = d(t_{ij}, t_{ik})$ is the distance between measurement times, $d_{\min} = \min_{i,j < k} d_{ijk}$ and $d_{\max} = \max_{i,j < k} d_{ijk}$ are computational constants equal to the minimum and maximum number of distance units across all subjects, respectively; ρ is the correlation parameter; and δ is the decay speed. We assume $0 \leq \rho < 1$ and $\delta \geq 0$ for reasons of practicality. The $d_{\min}$ and $d_{\max}$ constants allow the model to adapt to the data and scale distance such that the multiplier of the decay speed δ , $(d_{ijk} - d_{\min}) / (d_{\max} - d_{\min})$, is between 0 and 1 for computational purposes.

The LEAR correlation structure (3) includes several commonly used correlation models in longitudinal data analysis. When $\delta = 0$, this model reduces to the well-known compound symmetry correlation model. When $0 < \delta < d_{\max} - d_{\min}$, the correlation between measurements on a given subject decreases in time or space at a slower rate than that imposed by the AR(1) model. When $\delta = d_{\max} - d_{\min}$, this model reduces to the AR(1) correlation model. When $\delta > d_{\max} - d_{\min}$, the correlation between measurements on a given subject decreases in time at a faster rate than that imposed by the AR(1) model. As $\delta \rightarrow \infty$, this model approaches the moving average model of order 1, MA(1).

The full log-likelihood (ℓ_0) and weighted hybrid pairwise log-likelihood (ℓ_1), respectively, are as follows:

$$-\ell_0(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n \left[c_{0i} + c_{0i} \ln \sigma^2 + \frac{1}{2} K_{0i}(\rho, \delta) + \frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{B}_{0i} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right], \quad (4)$$

$$-\ell_1(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n \left[c_{1i} + c_{1i} \ln \sigma^2 + \frac{1}{2} K_{1i}(\rho, \delta) + \frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{B}_{1i} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right], \quad (5)$$

where $c_0 = c_{0i} \ln(2\pi)$ and $c_1 = c_{1i} \ln(2\pi)$ could be omitted during further calculations, $c_{0i} = m_i/2$, $c_{1i} = \frac{1}{2} \sum_j \sum_k w_{ijk}$, $K_{0i}(\rho, \delta) = \ln |\mathbf{R}_i|$, $\mathbf{B}_{0i} = \mathbf{R}_i^{-1}$, and

$$\begin{aligned} \mathbf{B}_{1i} &= \sum_{j < k} w_{ijk} \mathbf{A}_{i,jk}^\top (\mathbf{A}_{i,jk} \mathbf{R}_i \mathbf{A}_{i,jk}^\top)^{-1} \mathbf{A}_{i,jk} + \sum_j w_{ijj} \mathbf{E}_j \mathbf{E}_j^\top = \\ &= \sum_{j < k} w_{ijk} \left(\frac{1}{1 - \rho_{ijk}^2} (\mathbf{E}_j \mathbf{E}_j^\top + \mathbf{E}_k \mathbf{E}_k^\top) - \frac{\rho_{ijk}}{1 - \rho_{ijk}^2} (\mathbf{E}_j \mathbf{E}_k^\top + \mathbf{E}_k \mathbf{E}_j^\top) \right) + \\ &= \sum_j w_{ijj} \mathbf{E}_j \mathbf{E}_j^\top = \\ &= \sum_j \left(w_{ijj} \mathbf{E}_j \mathbf{E}_j^\top + \sum_{k \neq j} w_{ijk} \left(\frac{1}{1 - \rho_{ijk}^2} \mathbf{E}_j \mathbf{E}_j^\top - \frac{\rho_{ijk}}{1 - \rho_{ijk}^2} \mathbf{E}_j \mathbf{E}_k^\top \right) \right), \\ K_{1i}(\rho, \delta) &= \sum_{j < k} w_{ijk} \ln(1 - \rho_{ijk}^2), \end{aligned} \quad (6)$$

where $\rho_{ijk} = \rho^{d_{\min} + \delta(d_{ijk} - d_{\min}) / (d_{\max} - d_{\min})}$. The $2 \times m_i$ matrix $\mathbf{A}_{i,jk} = (\mathbf{E}_j, \mathbf{E}_k)^\top$ with $\mathbf{E}_j$ being the j th unit vector whose j th component is one and all others zero. In general, there are no simpler forms for $|\mathbf{R}_i|$ and $\mathbf{R}_i^{-1}$. Note that ρ and δ , but neither $\boldsymbol{\beta}$ nor σ^2 , appears in the $\mathbf{B}_{0i}$ and $\mathbf{B}_{1i}$ matrices.

2.1 Estimation of $\boldsymbol{\beta}$ and σ^2 with ρ and δ known

From Eqs. (4) and (5), we have the maximum likelihood estimators (MLE) and weighted hybrid pairwise likelihood estimators (WPLE) for the regression coefficient $\boldsymbol{\beta}$ as follows, with ρ and δ known:

$$\hat{\boldsymbol{\beta}}_l = \left[\sum_i \mathbf{X}_i^\top \mathbf{B}_{li} \mathbf{X}_i \right]^{-1} \sum_i \mathbf{X}_i^\top \mathbf{B}_{li} \mathbf{Y}_i, \quad l = 0, 1. \quad (7)$$

And the variances of (7) are

$$\text{Var}(\hat{\boldsymbol{\beta}}_l) = \sigma^2 \left[\sum_i \mathbf{X}_i^\top \mathbf{B}_{li} \mathbf{X}_i \right]^{-1} \sum_i \mathbf{X}_i^\top \mathbf{B}_{li} \mathbf{R}_i \mathbf{B}_{li} \mathbf{X}_i \left[\sum_i \mathbf{X}_i^\top \mathbf{B}_{li} \mathbf{X}_i \right]^{-1}. \quad (8)$$

Correspondingly, the MLE and WPLE for σ^2 are respectively, with $\boldsymbol{\beta}$, ρ , and δ assumed to be known,

$$\hat{\sigma}_l^2 = \frac{\sum_i (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_l)^\top \mathbf{B}_{li} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_l)}{2 \sum_i c_{li}}, \quad l = 0, 1. \quad (9)$$

And the variances of (9) are

$$\text{Var}(\hat{\sigma}_l^2) = \frac{\sigma^4}{2 \left[\sum_i c_{li} \right]^2} \sum_i \text{tr}([\mathbf{B}_{li} \mathbf{R}_i]^2). \quad (10)$$

For the optimal WPLE, we have the following theorem.

Theorem 1. With ρ and δ known, there exists a set of optimal weights $\{w_{ijk}^0\}$ such that the full log-likelihood (ℓ_0) and weighted hybrid pairwise log-likelihood (ℓ_1) differ by a constant across all $\boldsymbol{\beta}$ and σ^2 . In particular, the optimal weights are

$$w_{ijk}^0 = -\frac{(1 - \rho_{ijk}^2) \rho_{ijk}^-}{\rho_{ijk}}, \quad \text{for } j \neq k; \quad w_{ijj}^0 = \sum_k \frac{\rho_{ijk}^-}{\rho_{ijk}}, \quad (11)$$

where $\rho_{ijk}^- = \mathbf{B}_{0i,jk}$ and $\rho_{ijk} = \mathbf{R}_{i,jk}$.

Proof. Since

$$\ell_1 - \ell_0 = \sum_i \left[(c_{0i} - c_{1i}) \ln \sigma^2 + \frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top (\mathbf{B}_{0i} - \mathbf{B}_{1i}) (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] + C,$$

to make $\ell_1 - \ell_0$ a constant for all $\boldsymbol{\beta}$ and σ^2 , we must have $\mathbf{B}_{0i} = \mathbf{B}_{1i}$ for all i . Using (6), $\rho_{ijk}^- = \mathbf{B}_{0i,jk} = \mathbf{B}_{1i,jk} = -\frac{w_{ijk} \rho_{ijk}}{1 - \rho_{ijk}^2}$, which leads to

$$w_{ijk} = -\frac{(1 - \rho_{ijk}^2) \rho_{ijk}^-}{\rho_{ijk}}.$$

Note that $\rho_{ijj}^- = \mathbf{B}_{1i,jj} = \sum_{k \neq j} \frac{w_{ijk}}{1 - \rho_{ijk}^2} + w_{ijj}$, we have

$$w_{ijj} = -\sum_{k \neq j} \frac{w_{ijk}}{1 - \rho_{ijk}^2} + \rho_{ijj}^- = \sum_k \frac{\rho_{ijk}^-}{\rho_{ijk}}.$$

When $\mathbf{B}_{0i} = \mathbf{B}_{1i}$, it is easy to show that

$$\begin{aligned} c_{1i} &= -\sum_{j < k} \frac{(1 - \rho_{ijk}^2) \rho_{ijk}^-}{\rho_{ijk}} + \frac{1}{2} \sum_j \sum_k \frac{\rho_{ijk}^-}{\rho_{ijk}} = \\ &= \frac{1}{2} \left(\sum_j \rho_{ijj}^- + \sum_{j \neq k} \left(\frac{\rho_{ijk}^-}{\rho_{ijk}} - \frac{(1 - \rho_{ijk}^2) \rho_{ijk}^-}{\rho_{ijk}} \right) \right) = \\ &= \frac{1}{2} \left(\sum_j \rho_{ijj}^- + \sum_{j \neq k} \rho_{ijk} \rho_{ijk}^- \right) = \frac{1}{2} \left(\sum_j \sum_k \rho_{ijk} \rho_{ijk}^- \right) = \\ &= \frac{1}{2} \text{tr}(\mathbf{R}_i \mathbf{R}_i^{-1}) = \frac{m_i}{2} = c_{0i}. \end{aligned}$$

Therefore, both $\mathbf{B}_{0i} = \mathbf{B}_{1i}$ and $c_{0i} = c_{1i}$ hold under the optimal weights, i.e., the weighted pairwise likelihood yields the same result as the full likelihood (up to a constant).

2.2 Simultaneous estimation of $\boldsymbol{\beta}$, σ^2 , ρ , and δ

When $\boldsymbol{\beta}$, σ^2 , ρ , and δ are simultaneously estimated by (2), asymptotic normality holds for the weighted hybrid pairwise likelihood estimators. The inverse Godambe matrix for the weighted hybrid pairwise likelihood estimate is

$$\mathbf{G}^{-1} = \left(\sum_i \mathbf{H}_{li} \right)^{-1} \left(\sum_i \mathbf{V}_{li} \right) \left(\sum_i \mathbf{H}_{li} \right)^{-1}.$$

Note that the log-likelihood (4) and weighted hybrid pairwise log-likelihood (5) have similar forms, and the first-order partial derivatives of the log-likelihood functions for $l = 0, 1$ are as follows:

$$\begin{aligned}\frac{\partial \ell_l}{\partial \beta} &= \frac{1}{\sigma^2} \sum X^T B_{li}(\rho, \delta)(Y_i - X\beta), \\ \frac{\partial \ell_l}{\partial \sigma^2} &= \sum_i \left[-\frac{c_{li}}{\sigma^2} + \frac{1}{2} \frac{1}{\sigma^4} (Y_i - X\beta)^T B_{li}(\rho, \delta)(Y_i - X\beta) \right], \\ \frac{\partial \ell_l}{\partial \rho} &= \sum_i \left[-\frac{\partial K_{li}(\rho, \delta)}{2\partial \rho} - \frac{1}{2} \frac{1}{\sigma^2} (Y_i - X\beta)^T \frac{\partial B_{li}(\rho, \delta)}{\partial \rho} (Y_i - X\beta) \right], \\ \frac{\partial \ell_l}{\partial \delta} &= \sum_i \left[-\frac{\partial K_{li}(\rho, \delta)}{2\partial \delta} - \frac{1}{2} \frac{1}{\sigma^2} (Y_i - X\beta)^T \frac{\partial B_{li}(\rho, \delta)}{\partial \delta} (Y_i - X\beta) \right].\end{aligned}$$

The V_{li} matrix comes from the covariance of the above first-order derivatives, and the H_{li} matrix has the expected values of the negative Hessian of ℓ_l . The entries of the H_{li} and V_{li} matrices are shown in Table 1.

Unfortunately, when β , σ^2 , ρ , and δ are estimated simultaneously, there are no explicit forms in general for the estimators. Additionally, no weights exist to make both $K_{0i} = K_{1i}$ and $B_{0i} = B_{1i}$ for $m_i > 2$, which suggests that WPLeS are not as efficient as MLEs.

Lemma 1. With the weights provided in (11) and $m_i > 2$, $K_{0i}(\rho, \delta) \neq K_{1i}(\rho, \delta)$ holds almost everywhere.

Proof. Intuitively, K_{1i} is not always equal to K_{0i} , but the proof is tricky since the calculation of K_{1i} involves a matrix inverse. To avoid massive calculations, we introduce complex analysis, using the isolatedness theorem of the zero points in analytic functions to prove this lemma. The proof is threefold: First, prove that $\exp(K_{0i})$ and $\exp(K_{1i})$ are analytic functions; second, prove that for any $m_i > 2$, K_{0i} may not always equal to K_{1i} ; finally, with the isolatedness theorem of the zero points in analytic functions and the Fubini theorem, we prove that $K_{0i} \neq K_{1i}$ (a.e.).

Notice that,

$$R_{ijk} = \exp\left(\left(d_{\min} + \delta \frac{d_{ijk} - d_{\min}}{d_{\max} - d_{\min}}\right) \ln \rho\right),$$

thus $|R|$ is analytic on $(\rho, \delta) \in D_0 := (\mathbb{C} \setminus (-\infty, 0]) \times \mathbb{C}$. Since $w_{ijk} = -\frac{(1 - \rho_{ijk}^2)\rho_{ijk}^-}{\rho_{ijk}}$ and ρ_{ijk}^- in (11) can be represented as the adjoint matrix of R_i times $|R_i|$; thus, $\exp(K_{1i}(\rho, \delta)) := \prod_{j < k} (1 - \rho_{ijk}^2)^{w_{ijk}}$ is analytic on

$$\begin{aligned}(\rho, \delta) \in D_1 &:= [(\mathbb{C} \setminus (-\infty, 0] \times \mathbb{C}) \cap \{(\rho, \delta) \in \mathbb{C}^2 \mid |R_i| \neq 0\}] \cap \\ &\{(\rho, \delta) \in \mathbb{C}^2 \mid \sum_{j < k} I(\rho_{ijk} \in (-\infty, 0]) = 0\}.\end{aligned}$$

Thus $\exp(K_{0i}(\rho, \delta)) - \exp(K_{1i}(\rho, \delta))$ is analytic on $D_0 \cap D_1$.

Notice that, with $\delta = 0$,

$$w_{ijk} = \frac{1 - \rho_{ijk}^2}{\rho_{ijk}} \frac{\rho_{ijk}}{(1 - \rho_{ijk})(1 + (m_i - 1)\rho_{ijk})} = \frac{1 + \rho_{ijk}}{1 + (m_i - 1)\rho_{ijk}}.$$

Thus with $\rho = 1/2$, $\delta = 0$, $\exp(K_{0i}(1/2, 0)) = |R_i| = (m + 1)/2^m$, and $\exp(K_{1i}(\rho, \delta)) = (3/4)^{\frac{m(m-1)}{2} \frac{3}{1+m}}$. Since

$$f(m) = \left(\frac{3(m-1)m}{m+1} - m\right) \ln(2) + \ln(m+1) - \frac{3(m-1)m \ln(3)}{2(m+1)}$$

is monotonically decreasing with $m > 2$, and $f(2) = 0$, we have $f(m) < f(2) = 0$, which means that, for all $m \geq 3$,

$$\frac{1+m}{2^m} < \left(\frac{3}{4}\right)^{\frac{3m(m-1)}{2(m+1)}}.$$

In case $m_i > 2$, notice that $K_{0i}(\rho, \delta) - K_{1i}(\rho, \delta)$ is analytic, the set $\{(\rho, \delta) : K_{0i}(\rho, \delta) - K_{1i}(\rho, \delta) \neq 0\}$ is an open set, and thus is Lebesgue measurable, which suggests the set $\{(\rho, \delta) \mid K_{0i}(\rho, \delta) - K_{1i}(\rho, \delta) = 0\}$ is also Lebesgue measurable. Since $K_{0i}(\frac{1}{2}, 0) \neq K_{1i}(\frac{1}{2}, 0)$, with the isolatedness theorem of the zero points in analytic functions, for almost all fixed $\rho = \rho^*$, the zeroes of $d_{\rho^*}(\delta) = K_{0i}(\rho^*, \delta) - K_{1i}(\rho^*, \delta)$ is at most denumerable. Thus with Fubini theorem,

Table 1. The entries of H_{li} and V_{li} matrices.

Element	H_{li}	V_{li}
(β, β)	$\sigma^{-2} X^T B_{li}(\rho, \delta) X$	$\sigma^{-2} X^T B_{li}(\rho, \delta) R_i(\rho, \delta) B_{li}(\rho, \delta) X$
(β, σ^2)	0	0
(β, ρ)	0	0
(β, δ)	0	0
(σ^2, σ^2)	$-c_{li} \sigma^{-4} + \sigma^{-4} \text{tr}(B_{li}(\rho, \delta) R_i(\rho, \delta))$	$\frac{1}{2} \sigma^{-4} \text{tr}(B_{li}(\rho, \delta) R_i(\rho, \delta) B_{li}(\rho, \delta) R_i(\rho, \delta))$
(σ^2, ρ)	$-\frac{1}{2} \sigma^{-2} \text{tr}\left(\frac{\partial B_{li}(\rho, \delta)}{\partial \rho} R_i(\rho, \delta)\right)$	$-\frac{1}{2} \sigma^{-2} \text{tr}\left(B_{li}(\rho, \delta) R_i(\rho, \delta) \frac{\partial B_{li}(\rho, \delta)}{\partial \rho} R_i(\rho, \delta)\right)$
(σ^2, δ)	$-\frac{1}{2} \sigma^{-2} \text{tr}\left(\frac{\partial B_{li}(\rho, \delta)}{\partial \delta} R_i(\rho, \delta)\right)$	$-\frac{1}{2} \sigma^{-2} \text{tr}\left(B_{li}(\rho, \delta) R_i(\rho, \delta) \frac{\partial B_{li}(\rho, \delta)}{\partial \delta} R_i(\rho, \delta)\right)$
(ρ, ρ)	$\frac{\partial^2 K_{li}(\rho, \delta)}{2\partial \rho^2} + \frac{1}{2} \text{tr}\left(\frac{\partial^2 B_{li}(\rho, \delta)}{\partial \rho^2} R_i(\rho, \delta)\right)$	$\frac{1}{2} \text{tr}\left(\frac{\partial B_{li}(\rho, \delta)}{\partial \rho} R_i(\rho, \delta) \frac{\partial B_{li}(\rho, \delta)}{\partial \rho} R_i(\rho, \delta)\right)$
(ρ, δ)	$\frac{\partial^2 K_{li}(\rho, \delta)}{2\partial \rho \partial \delta} + \frac{1}{2} \text{tr}\left(\frac{\partial^2 B_{li}(\rho, \delta)}{\partial \rho \partial \delta} R_i(\rho, \delta)\right)$	$\frac{1}{2} \text{tr}\left(\frac{\partial B_{li}(\rho, \delta)}{\partial \rho} R_i(\rho, \delta) \frac{\partial B_{li}(\rho, \delta)}{\partial \delta} R_i(\rho, \delta)\right)$
(δ, δ)	$\frac{\partial^2 K_{li}(\rho, \delta)}{2\partial \delta^2} + \frac{1}{2} \text{tr}\left(\frac{\partial^2 B_{li}(\rho, \delta)}{\partial \delta^2} R_i(\rho, \delta)\right)$	$\frac{1}{2} \text{tr}\left(\frac{\partial B_{li}(\rho, \delta)}{\partial \delta} R_i(\rho, \delta) \frac{\partial B_{li}(\rho, \delta)}{\partial \delta} R_i(\rho, \delta)\right)$

$$m(\{(\rho, \delta) | K_{0i}(\rho, \delta) - K_{1i}(\rho, \delta) = 0\}) = \int_0^1 \int_0^\infty I(K_{0i}(\rho, \delta) - K_{1i}(\rho, \delta) = 0) d\delta d\rho = 0,$$

which suggests with $m_i > 2$, $K_{0i}(\rho, \delta) \neq K_{1i}(\rho, \delta)$ holds almost everywhere.

Theorem 2. When β , σ^2 , ρ , and δ are estimated simultaneously, there are no weights such that WPLEs are as efficient as the MLEs whenever $m_i > 2$.

Proof. According to Theorem 1, we can easily determine that the weights that yield optimal estimation for β and σ^2 are unique; thus, the optimal weight, if it exists, is fixed. With Lemma 1 and the weights in (11), $K_{1i}(\rho, \delta) \neq K_{0i}(\rho, \delta)$ holds almost everywhere with $m_i > 2$; thus the estimations of ρ and δ are inefficient at the given weights. Therefore, there are no weights that yield efficient estimation given that R_i is unknown.

Remark 1. From the standard theory of estimating equations and (2), the maximum composite likelihood estimators of β , σ^2 , ρ , and δ are asymptotically unbiased, consistent, and asymptotically normally distributed^[6].

3 Simulation

In this section, we conduct several simulation studies under multivariate normal distribution with different LEAR structures and weighting schemes to investigate the bias, the root mean squared error (RMSE), and the asymptotic relative

efficiencies (AREs, the ratios of diagonal elements of the inverse Fisher information matrix (the asymptotic variance matrix of the maximum likelihood estimate) with the corresponding diagonal elements of the estimated covariance matrix).

Ind-weights: $w_{ijk} = 1/(m_i - 1)$. Suppose only pairwise weights are used, and data has both $\rho = 0$ and $\delta = 0$, then the full-likelihood equals to weighted pairwise likelihood with $w_{ijk} = \mathbb{I}(j < k)/(m_i - 1)$.

Ar1-weights: Suppose the data has AR(1) structure, thus full likelihood equals to weighted pairwise likelihood with $w_{ijk} = \mathbb{I}(j + 1 = k)$ for $j \neq k$ and $w_{ijj} = -\mathbb{I}(j! = 1)$.

Adj1-weights: just a simplification of ar1-weights that only uses $w_{ijk} = \mathbb{I}(j + 1 = k)$.

In all the studies, the number of repeated measurements, m_i , of the i th subject is uniformly sampled from $\{1, 2, \dots, 10\}$. The corresponding observation times t_{ij} are then sampled uniformly from $\{1, 2, \dots, 10\}$. The responses $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{im_i})^T$ are generated from $N(X_i\beta, \sigma^2 R_i(\rho, \delta))$, where $X_{ij} = (1, t_{ij}, u_{ij})$ with $u_{ij} \sim N(0, 1)$, $\beta = (1, -0.5, 0.5)^T$, $\sigma^2 = 3$, and R_i is the LEAR correlation structure (3). We consider the following three commonly used correlation structures:

- Compound symmetry: $\delta = 0$.
- Nearly compound symmetry (which is between compound symmetry and AR(1) structures): $\delta = 1$.
- AR(1): $\delta = d_{\max} - d_{\min} = 9$.

Table 2 reports the results under a compound symmetric correlation structure. All three weighting methods result in

Table 2. Compound symmetric correlation structure.

Weighting methods		$n = 50$			$n = 100$			$n = 200$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	0.004	0.239	1.000	0.000	0.162	1.000	0.002	0.121	1.000
	β_1	0.001	0.031	1.000	0.001	0.022	1.000	-0.000	0.016	1.000
	β_2	0.004	0.092	1.000	0.003	0.065	1.000	0.003	0.047	1.000
	σ^2	-0.054	0.334	1.000	-0.028	0.206	1.000	-0.006	0.159	1.000
	ρ	0.004	0.077	1.000	0.004	0.048	1.000	0.005	0.038	1.000
	δ	0.155	0.313	1.000	0.099	0.202	1.000	0.064	0.126	1.000
Ind-weights	β_0	0.002	0.243	0.984	-0.001	0.168	0.966	0.000	0.127	0.953
	β_1	0.001	0.032	0.987	0.001	0.022	0.989	0.000	0.016	0.993
	β_2	0.004	0.098	0.937	0.004	0.069	0.944	0.003	0.050	0.944
	σ^2	-0.070	0.314	1.075	-0.033	0.212	0.975	-0.013	0.156	1.021
	ρ	0.003	0.071	1.083	0.005	0.048	0.985	0.005	0.036	1.046
	δ	0.198	0.383	0.828	0.128	0.241	0.864	0.085	0.161	0.796
Ar1-weights	β_0	0.004	0.249	0.958	0.001	0.171	0.946	0.001	0.129	0.935
	β_1	0.001	0.034	0.935	0.001	0.024	0.937	0.000	0.017	0.942
	β_2	0.002	0.097	0.948	0.003	0.068	0.953	0.003	0.050	0.952
	σ^2	-0.069	0.311	1.086	-0.031	0.212	0.971	-0.013	0.157	1.016
	ρ	-0.000	0.074	1.029	0.002	0.058	0.826	0.001	0.040	0.932
	δ	1.736	4.460	0.066	0.947	2.716	0.069	0.409	0.843	0.148
Adj1-weights	β_0	-0.003	0.254	0.938	-0.004	0.178	0.910	-0.002	0.135	0.896
	β_1	0.002	0.034	0.916	0.002	0.024	0.926	0.000	0.017	0.931
	β_2	-0.000	0.102	0.905	0.002	0.072	0.902	0.003	0.052	0.898
	σ^2	-0.071	0.325	1.040	-0.034	0.220	0.937	-0.011	0.163	0.975
	ρ	-0.003	0.075	1.014	0.001	0.057	0.827	0.002	0.041	0.921
	δ	1.694	4.374	0.067	0.920	2.584	0.073	0.404	0.835	0.149

similar biases, which decrease as n increases, and all the estimates except that of δ are unbiased even under small sample sizes. The estimate of δ is biased, but the bias decreases as n increases. This indicates that, compared with the other parameters, δ is more sensitive to the weights. As all the parameters are estimated simultaneously, the estimates of regression coefficients β_0 , β_1 , and β_2 are less efficient than the MLE. Furthermore, note that the true $\rho = 0.3$ indicates a weak correlation, the performance of estimates with ind-weights is least favorable to the MLE. For the second case with a nearly compound symmetric correlation, the results in Table 3 show that the ind-weights perform closest to the MLE. For the third case with an AR(1) correlation structure, the results shown in Table 4 indicate that the estimates with ar1-weights are slightly different from the MLE, although the component likelihood for the AR(1) structure is the same as the full likelihood with $\delta = d_{\max} - d_{\min} = 9$ theoretically; here, δ is specified as an unknown parameter. When δ changes, the results yielded by the AR(1) structure are slightly different from the full likelihood's result. The ar1-weights yield greater relative

efficiency than the other two methods and is almost as efficient as the MLE in estimating regression coefficients.

We also investigated the performance of these three weighting methods with $n = 200$ fixed and ρ set to 0.3, 0.5, and 0.7. The results are reported in Tables A1–A3 in the Appendix. The biases of $\hat{\beta}$ s vary slightly as ρ increases, since estimates of β s are asymptotically unbiased even if the correlation structure is misspecified. In comparison, the estimate of δ has a larger bias and RMSE as ρ increases. Moreover, the RMSE of most parameter estimates becomes larger as the inner correlation increases, because the weighting methods fail to reflect the true correlation. For example, when the true correlation is a compound symmetry structure, the asymptotic efficiency is lower for ind-weights with a higher ρ since a higher ρ means a stronger relationship, which does not meet the independence requirement of ind-weights. As the ar1-weights and adj1-weights describe the (near) compound symmetric structure more poorly with increasing ρ , the asymptotic relative efficiency also decreases. Similar conclusions can be drawn for simulations with near compound symmetric and

Table 3. Near compound symmetric correlation structure.

Weighting methods		$n = 50$			$n = 100$			$n = 200$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	0.004	0.259	1.000	0.000	0.177	1.000	0.002	0.132	1.000
	β_1	0.001	0.038	1.000	0.001	0.027	1.000	−0.000	0.019	1.000
	β_2	0.003	0.095	1.000	0.003	0.067	1.000	0.003	0.048	1.000
	σ^2	−0.064	0.294	1.000	−0.028	0.197	1.000	−0.011	0.145	1.000
	ρ	−0.007	0.071	1.000	−0.004	0.054	1.000	−0.003	0.037	1.000
	δ	0.420	2.132	1.000	0.140	0.761	1.000	0.054	0.489	1.000
Ind-weights	β_0	0.002	0.260	0.996	−0.001	0.181	0.980	0.001	0.137	0.968
	β_1	0.001	0.038	0.996	0.001	0.027	0.996	0.000	0.019	0.995
	β_2	0.003	0.102	0.940	0.003	0.071	0.940	0.003	0.052	0.935
	σ^2	−0.067	0.293	1.005	−0.031	0.199	0.993	−0.012	0.147	0.988
	ρ	−0.008	0.078	0.912	−0.004	0.058	0.932	−0.002	0.039	0.933
	δ	0.324	1.312	1.645	0.150	0.808	0.942	0.069	0.529	0.926
Ar1-weights	β_0	0.005	0.263	0.986	0.001	0.181	0.980	0.002	0.136	0.971
	β_1	0.001	0.039	0.978	0.001	0.027	0.982	−0.000	0.019	0.982
	β_2	0.002	0.098	0.978	0.003	0.069	0.975	0.003	0.050	0.971
	σ^2	−0.064	0.293	1.005	−0.029	0.199	0.993	−0.012	0.146	0.992
	ρ	−0.004	0.075	0.944	−0.001	0.056	0.963	−0.002	0.038	0.969
	δ	2.340	6.405	0.351	1.209	3.955	0.199	0.415	1.527	0.330
Adj1-weights	β_0	−0.002	0.273	0.951	−0.004	0.191	0.926	−0.001	0.144	0.915
	β_1	0.002	0.041	0.943	0.001	0.028	0.946	0.000	0.020	0.945
	β_2	−0.000	0.102	0.937	0.001	0.072	0.929	0.003	0.053	0.921
	σ^2	−0.067	0.309	0.953	−0.032	0.207	0.954	−0.011	0.153	0.947
	ρ	−0.007	0.075	0.954	−0.003	0.056	0.964	−0.003	0.038	0.957
	δ	2.251	6.151	0.365	1.134	3.722	0.211	0.402	1.517	0.332

Table 4. AR(1) correlation structure.

Weighting methods		$n = 50$			$n = 100$			$n = 200$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	0.003	0.259	1.000	−0.000	0.178	1.000	−0.000	0.134	1.000
	β_1	0.001	0.041	1.000	0.001	0.029	1.000	0.000	0.021	1.000
	β_2	0.002	0.098	1.000	0.003	0.069	1.000	0.003	0.050	1.000
	σ^2	−0.052	0.282	1.000	−0.021	0.189	1.000	−0.007	0.140	1.000
	ρ	−0.005	0.067	1.000	−0.002	0.049	1.000	−0.002	0.034	1.000
	δ	3.943	10.629	1.000	3.011	8.313	1.000	1.697	5.895	1.000
Ind-weights	β_0	0.000	0.259	0.999	−0.002	0.180	0.990	−0.001	0.136	0.982
	β_1	0.001	0.041	0.995	0.001	0.029	0.995	0.000	0.021	0.992
	β_2	0.004	0.104	0.942	0.003	0.073	0.941	0.003	0.053	0.937
	σ^2	−0.055	0.280	1.007	−0.023	0.189	1.005	−0.007	0.140	0.997
	ρ	−0.014	0.082	0.821	−0.007	0.060	0.817	−0.002	0.039	0.865
	δ	1.745	6.947	1.468	1.346	5.721	1.394	1.197	4.699	1.242
Ar1-weights	β_0	0.005	0.259	1.002	0.000	0.178	1.002	0.000	0.134	0.998
	β_1	0.000	0.041	0.999	0.001	0.029	1.000	0.000	0.021	0.999
	β_2	0.002	0.097	1.005	0.004	0.069	1.001	0.003	0.050	0.999
	σ^2	−0.053	0.281	1.003	−0.022	0.189	1.003	−0.007	0.140	1.000
	ρ	−0.006	0.067	0.991	−0.003	0.050	0.972	−0.002	0.034	0.985
	δ	5.093	13.029	0.823	3.719	9.989	0.836	3.385	9.137	0.665
Adj1-weights	β_0	−0.003	0.273	0.949	−0.005	0.192	0.929	−0.003	0.144	0.927
	β_1	0.001	0.044	0.940	0.001	0.031	0.937	0.000	0.022	0.937
	β_2	0.001	0.102	0.962	0.002	0.072	0.956	0.003	0.053	0.946
	σ^2	−0.055	0.295	0.954	−0.023	0.195	0.972	−0.005	0.145	0.962
	ρ	−0.009	0.067	0.996	−0.005	0.050	0.977	−0.003	0.035	0.980
	δ	4.420	12.028	0.882	3.469	9.660	0.859	3.156	8.838	0.684

AR(1) correlation structures. Notably, the bias of δ is greater than that of the other parameters and tends to decrease as ρ increases. This is because δ measures the rate of correlation decline with respect to the time lag. When ρ is small, the change in δ will not be significant.

In summary, the ind-weights method performs as efficiently as the full likelihood when the repeated observations of each subject are independent. The adj1-weights method will work best when the observations are only correlated adjacently, i.e., the within-subject correlation structure is 1-banded exchangeable correlation. The ar1-weights method is the best when the within-subject correlation structure is known as AR(1). Since maximizing the weighted pairwise likelihood with ar1-weights is faster and yields consistent estimators, starting estimation with ar1-weights is a fruitful choice. After a consistent estimator is obtained, more accurate weights could be chosen to minimize the asymptotic variance of some specific parameter.

4 Analysis of CD4 data

In this section, we illustrate the weighted hybrid pairwise likelihood estimation approach via analysis of CD4 data. Patients in the AIDS Clinical Trial Group (ACTG) Study 193A were randomized to dual or triple combinations of HIV-1 reverse transcriptase inhibitors^[19]. In this study, a total of 1309 patients were randomly assigned to one of four daily regimens, each containing 600 mg of zidovudine. The four regimens were: monthly alternating zidovudine with 400 mg of didanosine, zidovudine plus 2.25 mg of zalcitabine, zidovudine plus 400 mg of didanosine, and triple therapy consisting of zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine. Measurements of CD4 counts were scheduled to be collected at baseline and at 8-week intervals during follow-up. However, the CD4 count data are unbalanced due to mistimed measurements and missing data resulting from skipped visits and dropouts. The number of measurements of CD4 counts during the first 40 weeks of follow-up varied

from 1 to 9, with a median of 4. The response variable is the log-transformed CD4 count, $\log(\text{CD4 count} + 1)$. The dataset is publicly available in the R package ALA.

To investigate the efficiency of different weighting methods, we consider the following regression model^[20]:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 (t_{ij} - 16)_+ + \beta_4 \text{Group}_i \times \{t_{ij} - (t_{ij} - 16)_+\} + \beta_5 \text{Age}_i + \beta_6 \text{Gender}_i + e_{ij},$$

where t_{ij} denotes the time since the baseline for the j th measurement on the i th subject (with $t_{i1} = 0$ at the baseline). $\text{Group}_i = 1$ if the i th subject is randomized to triple therapy, and $\text{Group}_i = 0$ otherwise; $(t_{ij} - 16)_+ = t_{ij} - 16$ if $t_{ij} > 16$ and $(t_{ij} - 16)_+ = 0$ if $t_{ij} \leq 16$. Age_i is the baseline age (in years) of the patients, $\text{Gender}_i = 1$ if the i th patient is male, and $\text{Gender}_i = 0$ otherwise. In this model, the mean rate of change from baseline to Week 16 may differ between the two groups (with slopes of β_2 and $\beta_2 + \beta_4$ respectively), but the mean rate of change from Week 16 to Week 40 is assumed to be the same (with a slope of $\beta_2 + \beta_3$).

To investigate the correlation structure, we first plot the smoothed autocorrelation and sample correlation matrix of the CD4 data as shown in Fig. 1a and Fig. 1b, respectively, where the smoothed autocorrelation at time t is calculated by $\text{scor}_d(t) = \text{corr}_{(i,j) \in T_d(t), (i,k) \in T_d(t)}(y_{ij}, y_{ik})$ with $T_d(t) = \{(i, j) : |t_{ij} - t| \leq d\}$ and $d = 3$ (days) being chosen to determine the size of the smoothing window. Sample correlation is calculated by first dividing the time into 9 regions and then calculating the correlation in each region. The heterogeneity in the first several months suggests that the model is not a stationary model and that the time effect should not be considered linearly, whereas the slightly decreasing trend of the smoothed correlation suggests a near-compound symmetric correlation structure. Notably, the linear regression of log correlation (logcor) and time lag (tlag) for smoothed autocorrelation yields

$$\text{logcor} = -0.2707 - 0.009393\text{tlag},$$

which directly leads to

$$\hat{\rho} = 0.3842, \hat{\delta} = 0.3899. \quad (12)$$

By fixing such estimates $\hat{\rho}$ and $\hat{\delta}$, the empirical optimal weights can be obtained via (11); thus, empirical optimal estimates for regression coefficients and the variance of error can be calculated via (7) and (9), respectively. To further assess effects of ρ and δ , we estimate the parameters under different values of ρ and δ . The results are reported in Table A4 in the Appendix. The estimated $(\hat{\rho}, \hat{\delta}) = (0.3842, 0.3899)$ yields the smallest negative likelihood value of 5990.799. The estimates of β s vary negligibly with changing ρ and δ , but $\hat{\sigma}^2$ changes slightly, as the estimates of β s are asymptotically unbiased even when ρ , δ , and σ^2 are misspecified.

We compare the estimates under the empirical optimal weights, the ar1 and ind weights, with the full likelihood estimates. The results are shown in Table 5. All four approaches yield similar results, with β_1 to β_5 being significant. The MLE approach has the smallest standard deviation. And the empirical optimal weights result in the most efficient estimates among the three weighting schemes, which is

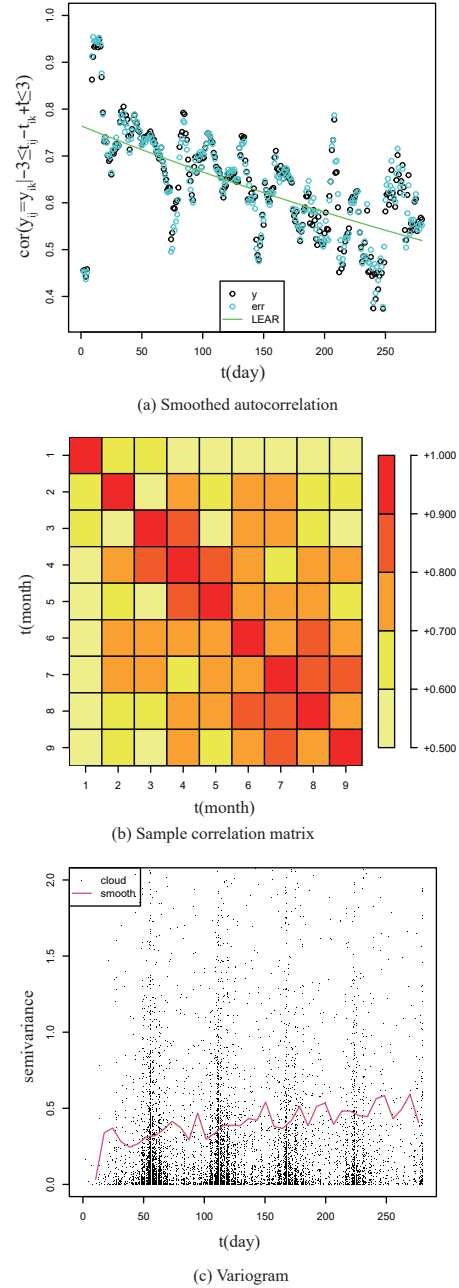


Fig. 1. Plots of the smoothed autocorrelation, empirical correlation matrix, and variogram of the CD4 data.

almost as efficient as the MLE. These findings are consistent with the effectiveness of triple therapy reported in Ref. [19], and the time effect is consistent with that mentioned in Ref. [20]. To further justify the near compound symmetric correlation structure, we calculate the residual $\hat{e}_{ij} = Y_{ij} - X_{ij}\hat{\beta}$ and define the smoothed semivariance

$$v_d(u) = \frac{\sum_{i, u \leq t_{i2} - t_{i1} \leq u+d} (\hat{e}_{i2} - \hat{e}_{i1})^2}{2 \sum_{i, u \leq t_{i2} - t_{i1} \leq u+d} 1},$$

where the size of the smoothing window is d . Fig. 1c shows both the variogram cloud and the smoothed variogram plot

Table 5. Estimate results.

	Full likelihood		Empirical optimal		Ind-weights		Ar1-weights	
	Est	SD	Est	SD	Est	SD	Est	SD
β_1	2.6170	0.1331	2.6170	0.1323	2.6333	0.1410	2.6183	0.1325
β_2	-0.0061	0.0018	-0.0060	0.0018	-0.0060	0.0019	0.0046	0.0019
β_3	-0.0128	0.0028	-0.0128	0.0028	-0.0125	0.0029	-0.0130	0.0030
β_4	0.0258	0.0030	0.0258	0.0030	0.0238	0.0034	0.0252	0.0032
β_5	0.0111	0.0031	0.0111	0.0031	0.0114	0.0033	0.0107	0.0031
β_6	-0.1142	0.0782	-0.1142	0.0778	-0.1169	0.0832	-0.1228	0.0779
σ^2	1.1085	0.0318	1.1007	0.0073*	1.1285	0.0342	1.0982	0.0524
ρ	0.3890	0.0162	0.3842	—	0.4100	0.0356	0.4052	0.0573
δ	0.3898	0.0413	0.3899	—	0.4210	0.0842	0.5008	0.0361
$-\ln l$	5990.7508		5990.7987		5992.0025		5992.2928	

* Report the theoretical asymptotic variance with assumption that $\rho = 0.3842$ and $\delta = 0.3899$.

generated by the smoothed semivariance that we calculate. The smoothed variogram increases with time lag, corresponding to decreasing correlation as residuals are separated in time.

5 Conclusions

We considered the choice of weights for the weighted hybrid pairwise likelihood approach. Some analytic results are obtained via the multivariate normal distribution with the LEAR correlation structure, which includes the commonly used AR(1), compound symmetry, and first-order moving average correlation structures in longitudinal data analysis. Through a combination of analytical and numerical examples, we provide some suggestions for a practically good choice of weights.

Notably, the selection of optimal weights in general is difficult. The specified weights could lead to a significant loss of efficiency when that the model assumptions are violated. For example, when the correlation is far from the LEAR structure could represent, the weights for the least squares estimate (7) may become highly inefficient. Moreover, the pairwise likelihood function still requires the computation of a sum over all pairs of variables and all observations, which may be prohibitive in high dimensions.

To alleviate the model misspecification issue, it might be better to use weighted estimating equation approaches rather than maximizing weighted pairwise likelihood. A set of optimal weights for different parameter components can then be further considered. For the computational burden of pairwise likelihood, one possible remedy is finding the enhancements of the randomized Bernoulli weighting method proposed in Ref. [8]. These extensions are challenging but deserve further exploration.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (12171450).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Feiyue Xie received his Ph.D. degree from the University of Science and Technology of China. His research mainly focuses on longitudinal data analysis.

Weiping Zhang is a Professor at the University of Science and Technology of China (USTC). He received his Ph.D. degree from USTC in 2005. His research mainly focuses on longitudinal data analysis and Bayesian analysis.

References

- [1] Tang C, Zhang W, Leng C. Discrete longitudinal data modeling with a mean-correlation regression approach. *Statistica Sinica*, **2019**, 29 (2): 853–876.
- [2] Zhang W, Li Y, Chen Y, et al. Parsimonious Gaussian copula modelling through constrained Cholesky decomposition for data with temporal dependence. *SCIENTIA SINICA Mathematica*, **2023**, 53 (5): 777–790. (in Chinese)
- [3] Lindsay B G. Composite likelihood methods. *Contemporary Mathematics*, **1988**, 80 (1): 221–239.
- [4] Zhao Y, Joe H. Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, **2005**, 33 (3): 335–356.
- [5] Heagerty P J, Zeger S L. Lorelogram: a regression approach to exploring dependence in longitudinal categorical response. *Journal of the American Statistical Association*, **1998**, 93 (441): 150–162.
- [6] Kuk A Y C, Nott D J. A pairwise likelihood approach to analyzing correlated binary data. *Statistics & Probability Letters*, **2000**, 47 (4): 329–335.
- [7] Varin C, Reid N, Firth D. An overview of composite likelihood methods. *Statistica Sinica*, **2011**, 21 (1): 5–42.
- [8] Mazo G, Karlis D, Rau A. A randomized pairwise likelihood method for complex statistical inferences. *Journal of the American Statistical Association*, **2024**, 119 (547): 2317–2327.
- [9] Ferrari D, Qian G, Hunter T. Parsimonious and efficient likelihood composition by Gibbs sampling. *Journal of Computational and Graphical Statistics*, **2016**, 25 (3): 935–953.

- [10] Joe H, Lee Y. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, **2009**, 100 (4): 670–685.
- [11] Varin C, Host G, Skare O. Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics & Data Analysis*, **2005**, 49 (4): 1173–1191.
- [12] Kuk A Y C. A hybrid pairwise likelihood method. *Biometrika*, **2007**, 94 (4): 939–952.
- [13] Lindsay B G, Yi G Y, Sun J. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, **2011**, 21 (1): 71–105.
- [14] Vasdekis V G S, Rizopoulos D, Moustaki I. Weighted pairwise likelihood estimation for a general class of random effects models. *Biostatistics*, **2014**, 15 (4): 677–689.
- [15] Papageorgiou I, Moustaki I. Sampling of pairs in pairwise likelihood estimation for latent variable models with categorical observed variables. *Statistics and Computing*, **2019**, 29 (2): 351–365.
- [16] Simpson S L, Edwards L J, Muller K E, et al. A linear exponent AR(1) family of correlation structures. *Statistics in Medicine*, **2010**, 29 (17): 1825–1838.
- [17] Yi G Y. Composite likelihood/pseudolikelihood. In: Wiley StatsRef: Statistics Reference Online. New York: Wiley, **2017**.
- [18] Varin C. On composite marginal likelihoods. *Advances in Statistical Analysis*, **2008**, 92 (1): 1–28.
- [19] Henry K, Erice A, Tierney C, et al. A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. *Journal of Acquired Immune Deficiency Syndromes & Human Retrovirology*, **1998**, 19 (4): 339–349.
- [20] Fitzmaurice G M, Laird N M, Ware J H. Applied Longitudinal Analysis. New York: Wiley, **2011**.

Appendix

A.1 Additional simulation results

We have conducted additional simulations with $n = 200$ fixed and increasing ρ from 0.3 to 0.7 to investigate the effects of the ind-weights, ar1-weights, and adj1-weights methods. The results are reported in Tables A1–A3.

Table A1. Results under a compound symmetric correlation structure with different ρ values.

Weighting methods		$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	0.002	0.121	1.000	0.003	0.122	1.000	0.004	0.122	1.000
	β_1	−0.000	0.016	1.000	−0.000	0.013	1.000	−0.000	0.011	1.000
	β_2	0.003	0.047	1.000	0.002	0.040	1.000	0.002	0.031	1.000
	σ^2	−0.006	0.159	1.000	−0.006	0.198	1.000	−0.023	0.237	1.000
	ρ	0.005	0.037	1.000	0.002	0.037	1.000	−0.002	0.029	1.000
	δ	0.064	0.126	1.000	0.046	0.089	1.000	0.037	0.071	1.000
Ind-weights	β_0	0.000	0.127	0.953	0.002	0.132	0.923	0.003	0.135	0.900
	β_1	0.000	0.016	0.993	0.000	0.014	0.987	0.000	0.011	0.987
	β_2	0.003	0.050	0.944	0.003	0.043	0.929	0.002	0.034	0.935
	σ^2	−0.013	0.156	1.017	−0.017	0.195	1.021	−0.019	0.239	0.993
	ρ	0.005	0.036	1.041	0.002	0.037	1.009	0.000	0.026	1.109
	δ	0.085	0.161	0.796	0.065	0.122	0.740	0.055	0.102	0.704
Ar1-weights	β_0	0.001	0.129	0.935	0.002	0.136	0.896	0.003	0.138	0.881
	β_1	0.000	0.017	0.942	0.000	0.016	0.861	0.000	0.014	0.756
	β_2	0.003	0.050	0.952	0.002	0.044	0.922	0.002	0.035	0.894
	σ^2	−0.013	0.157	1.013	−0.018	0.186	1.070	−0.016	0.231	1.023
	ρ	0.001	0.040	0.927	−0.001	0.037	1.005	−0.001	0.028	1.057
	δ	0.409	0.843	0.148	0.275	0.512	0.176	0.220	0.397	0.184
Adj1-weights	β_0	−0.002	0.135	0.896	−0.001	0.143	0.850	0.001	0.149	0.816
	β_1	0.000	0.017	0.931	0.000	0.015	0.881	0.000	0.013	0.805
	β_2	0.003	0.052	0.898	0.003	0.046	0.868	0.002	0.037	0.850
	σ^2	−0.011	0.163	0.973	−0.017	0.197	1.009	−0.021	0.253	0.936
	ρ	0.001	0.040	0.921	−0.001	0.035	1.044	−0.002	0.030	0.981
	δ	0.404	0.835	0.149	0.270	0.505	0.178	0.217	0.392	0.186

Table A2. Results under the near compound symmetric correlation structure with different ρ values.

Weighting methods		$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	0.002	0.132	1.000	0.004	0.134	1.000	0.005	0.131	1.000
	β_1	-0.000	0.019	1.000	-0.000	0.018	1.000	-0.000	0.015	1.000
	β_2	0.003	0.048	1.000	0.002	0.042	1.000	0.002	0.033	1.000
	σ^2	-0.011	0.145	1.000	-0.013	0.170	1.000	-0.015	0.211	1.000
	ρ	-0.003	0.037	1.000	-0.004	0.034	1.000	-0.003	0.026	1.000
	δ	0.054	0.489	1.000	0.016	0.321	1.000	0.001	0.255	1.000
Ind-weights	β_0	0.001	0.137	0.968	0.002	0.143	0.938	0.003	0.144	0.911
	β_1	0.000	0.019	0.995	-0.000	0.018	0.990	-0.000	0.015	0.989
	β_2	0.003	0.052	0.935	0.003	0.046	0.900	0.002	0.037	0.888
	σ^2	-0.012	0.147	0.988	-0.017	0.176	0.966	-0.020	0.226	0.938
	ρ	-0.002	0.039	0.933	-0.004	0.036	0.935	-0.004	0.028	0.925
	δ	0.069	0.529	0.926	0.027	0.365	0.882	0.011	0.297	0.860
Ar1-weights	β_0	0.002	0.136	0.971	0.003	0.142	0.945	0.004	0.142	0.925
	β_1	-0.000	0.019	0.982	-0.000	0.019	0.951	-0.000	0.017	0.899
	β_2	0.003	0.050	0.971	0.002	0.044	0.950	0.002	0.035	0.931
	σ^2	-0.012	0.146	0.993	-0.016	0.174	0.982	-0.018	0.217	0.976
	ρ	-0.002	0.038	0.969	-0.004	0.036	0.954	-0.003	0.027	0.949
	δ	0.415	1.527	0.330	0.158	0.912	0.357	0.081	0.723	0.355
Adj1-weights	β_0	-0.001	0.144	0.915	-0.000	0.152	0.880	0.001	0.155	0.846
	β_1	0.000	0.020	0.945	0.000	0.019	0.929	0.000	0.017	0.906
	β_2	0.003	0.053	0.921	0.003	0.047	0.895	0.002	0.037	0.882
	σ^2	-0.011	0.153	0.947	-0.017	0.185	0.922	-0.021	0.236	0.895
	ρ	-0.003	0.038	0.957	-0.004	0.036	0.936	-0.004	0.029	0.902
	δ	0.402	1.517	0.332	0.147	0.906	0.359	0.075	0.718	0.357

Table A3. Results under the AR(1) correlation structure with different ρ values.

Weighting methods		$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
		Bias	RMSE	ARE	Bias	RMSE	ARE	Bias	RMSE	ARE
Full likelihood	β_0	-0.000	0.134	1.000	0.001	0.143	1.000	0.003	0.149	1.000
	β_1	0.000	0.021	1.000	0.000	0.022	1.000	-0.000	0.022	1.000
	β_2	0.003	0.050	1.000	0.003	0.044	1.000	0.002	0.034	1.000
	σ^2	-0.007	0.140	1.000	-0.011	0.149	1.000	-0.015	0.169	1.000
	ρ	-0.002	0.034	1.000	-0.003	0.028	1.000	-0.002	0.020	1.000
	δ	1.697	5.895	1.000	0.300	1.866	1.000	0.104	1.155	1.000
Ind-weights	β_0	-0.001	0.136	0.982	0.000	0.151	0.951	0.002	0.163	0.913
	β_1	0.000	0.021	0.992	-0.000	0.023	0.968	-0.000	0.023	0.942
	β_2	0.003	0.053	0.937	0.003	0.051	0.858	0.002	0.044	0.769
	σ^2	-0.007	0.140	0.997	-0.011	0.151	0.987	-0.017	0.176	0.963
	ρ	-0.002	0.039	0.865	-0.002	0.033	0.849	-0.002	0.024	0.840
	δ	1.197	4.699	1.242	0.462	2.616	0.715	0.207	1.809	0.640
Ar1-weights	β_0	0.000	0.134	0.998	0.002	0.144	0.998	0.004	0.150	0.998
	β_1	0.000	0.021	0.999	-0.000	0.022	0.998	-0.000	0.022	0.997
	β_2	0.003	0.050	0.999	0.003	0.044	0.999	0.002	0.034	0.998
	σ^2	-0.007	0.140	1.000	-0.010	0.149	0.999	-0.015	0.169	1.000
	ρ	-0.002	0.034	0.985	-0.002	0.029	0.962	-0.002	0.021	0.951
	δ	3.385	9.137	0.665	1.151	4.551	0.418	0.388	2.595	0.448
Adj1-weights	β_0	-0.003	0.144	0.927	-0.002	0.160	0.898	0.001	0.171	0.870
	β_1	0.000	0.022	0.937	0.000	0.024	0.915	-0.000	0.024	0.905
	β_2	0.003	0.053	0.946	0.003	0.047	0.930	0.002	0.037	0.920
	σ^2	-0.005	0.145	0.962	-0.009	0.157	0.949	-0.016	0.184	0.917
	ρ	-0.003	0.035	0.980	-0.003	0.029	0.951	-0.003	0.022	0.916
	δ	3.156	8.838	0.684	1.080	4.468	0.425	0.343	2.572	0.451

A.2 Estimations with specified ρ and δ

We fixed ρ and δ to other values to show the effect on the estimation of the other parameters when ρ and δ are incorrectly specified. The results are reported in Table A4.

Table A4. Estimations with specified ρ and δ .

	Est	SD	Est	SD	Est	SD	Est	SD
β_1	2.6171	0.1324	2.6085	0.0823	2.6083	0.1583	2.6216	0.1147
β_2	-0.0061	0.0018	-0.0034	0.0028	-0.0069	0.0017	-0.0052	0.0020
β_3	-0.0128	0.0028	-0.0126	0.0046	-0.0126	0.0027	-0.0127	0.0031
β_4	0.0258	0.0030	0.0212	0.0028	0.0256	0.0027	0.0251	0.0031
β_5	0.0111	0.0031	0.0115	0.0018	0.0113	0.0037	0.0110	0.0027
β_6	-0.1142	0.0778	-0.1234	0.0478	-0.1066	0.0931	-0.1202	0.0674
σ^2	1.1008	0.0073	1.1270	0.0075	1.4373	0.0095	0.9426	0.0063
ρ	0.3842	—	0.0000	—	0.3842	—	0.3842	—
δ	0.3899	—	—	—	0.0000	—	1.0000	—
$-\ln l$	5990.7987		7446.7602		6089.8666		6049.8161	
β_1	2.6166	0.1294	2.6173	0.1274	2.6164	0.1364	2.6171	0.1342
β_2	-0.0061	0.0018	-0.0060	0.0018	-0.0062	0.0018	-0.0061	0.0018
β_3	-0.0128	0.0028	-0.0128	0.0029	-0.0128	0.0028	-0.0128	0.0028
β_4	0.0257	0.0030	0.0256	0.0030	0.0259	0.0030	0.0258	0.0030
β_5	0.0111	0.0030	0.0111	0.0030	0.0111	0.0032	0.0111	0.0031
β_6	-0.1141	0.0761	-0.1148	0.0749	-0.1134	0.0802	-0.1142	0.0789
σ^2	1.0660	0.0071	1.0456	0.0069	1.1464	0.0076	1.1220	0.0075
ρ	0.3500	—	0.3500	—	0.4000	—	0.4000	—
δ	0.3500	—	0.4000	—	0.3500	—	0.4000	—
$-\ln l$	5992.6455		5994.4366		5991.9383		5990.9084	