Journal of the Korean Data & Information Science Society
2016, **27**(3), 815–825

http://dx.doi.org/10.7465/jkdi.2016.27.3.815
한국데이터정보과학회지

Modeling of random effects covariance matrix in marginalized random effects models[†]

Keunbaik Lee[1] · Seolhwa Kim[2]

[12]Department of Statistics, Sungkyunkwan University

Received 12 February 2016, revised 7 March 2016, accepted 12 April 2016

Abstract

Marginalized random effects models (MREMs) are often used to analyze longitudinal categorical data. The models permit direct estimation of marginal mean parameters and specify the serial correlation of longitudinal categorical data via the random effects. However, it is not easy to estimate the random effects covariance matrix in the MREMs because the matrix is high-dimensional and must be positive-definite. To solve these restrictions, we introduce two modeling approaches of the random effects covariance matrix: partial autocorrelation and the modified Cholesky decomposition. These proposed methods are illustrated with the real data from Korean genomic epidemiology study.

Keywords: Autocorrelation, modified Cholesky decomposition, heterogeneity, Quasi-Monte Carlo.

1. Introduction

In longitudinal studies, outcomes are collected over time from same subjects. Therefore, the repeated outcomes are correlated and serial correlation must be taken into account to estimate the effect of covariate (Fitzmaurice and Laird, 1993). To explain the correlation, random effects are commonly used in likelihood-based modeling such as linear mixed models (Choi and Huh, 2014), generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993), and marginalized models (Heagerty, 1999; 2002).

In analysis of longitudinal categorical data, two likelihood-based models, GLMMs and marginalized random effects models (MREMs) (Heagerty, 1999), have been proposed. The GLMMs are preferable when subject-specific effects are of interest, and the MREMs are commonly used when population-average effect are of interest (Agresti, 2002). In these models, the serial correlation are explained via random effects and it is a well known restriction that random effects covariance matrix (RECM) in these models is high-dimensional and must be positive-definite. To remove these restriction it is often assumed that the matrix is constant

[†] This project was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A2054997).

[1] Corresponding author: Associate professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea. E-mail: keunbaik@skku.edu

[2] Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.

over subjects and is simply structured such as AR(1). These assumptions, however, are too strong in many situations and may result in serious biases for estimates of the fixed effects in GLMMs (Heagerty and Kurland, 2001). In this paper we consider general structures of the RECM to release the strong assumptions.

There are many methods to construct a general structures of (random effects) covariance matrix (Jeon and Lee, 2014). Especially, we focus on two recently developed methods: modified Cholesky decomposition and partial autocorrelation approaches. Both approaches reduce the number of parameters of the covariance matrix dramatically and the approaches provide an unconstrained and statistically interpretable reparameterization of the covariance matrix. The modified Cholesky decomposition approach decomposes the covariance matrix into two set of parameters, generalized autoregressive parameters (GARPs) and innovation variances (IVs) (Pourahmadi, 1999; 2000). The GARPs characterize the dependence structure and the IVs are a set of variance parameters. The GARP/IV parameters are natural for characterizing missingness due to dropout and the covariance matrix is positive definite when the IVs are positive (Pourahmadi, 1999; 2000). Bayesian modeling for the GARP and IV parameters were proposed in linear mixed models (Daniels and Pourahmadi, 2002; Pourahmadi and Daniels, 2002). Daniels and Zhao (2003) proposed a Bayesian modeling for the random effects covariance matrices in linear mixed models. Pan and MacKenzie (2003, 2007) used the modified Cholesky decomposition to estimate joint mean-covariance modeling for linear mixed models, respectively. Lee *et al.* (2012) proposed a Bayesian modeling for GLMMs using the modified Cholesky decomposition. Lee and Sung (2014) used the modified Cholesky decomposition to model RECM in MREM.

The (random effects) covariance matrix can be expressed using the diagonal matrix of standard deviations and the correlation matrix between responses over time. The correlation matrix has diagonal entries 1 and must be positive definite with elements taking values between -1 and 1. The MCD of the correlation matix can hardly satisfy the requirements. Therefore, alternative approach for modeling of the correlation matrix is the partial autocorrelation approach. The approach uses a one-to-one correspondence between a correlation matrix and its associated matrix of partial autocorrelations (Daniels and Pourahmadi, 2009). Since there is no restriction of the partial autocorrelation matrix such as positive-definiteness, generalized linear models are used for modeling of the partial autocorrelation matrix and the corresponding correlation matrix is guaranteed to be positive definite. Wang and Daniels (2013) proposed Bayesian modeling for the partial autocorrelations and discussed priors for the parameters of the partial autocorrelation matrix. Lee *et al.* (2013) extended this approach to accommodate bivariate longitudinal ordinal data. In this paper, we compare the two approaches for RECM in MREMs.

This paper is organized as follows. In Section 2, we review MREMs for longitudinal binary data using the modified Cholesky decomposition. In section 3, we propose the partial autocorrelation approach for the RECM in MREMs. The algorithm for the maximum likelihood estimation is proposed which uses the quasi-Newton algorithm. In section 4, we present analysis of longitudinal data on metabolic syndrome. Finally, we present conclusions and extensions in section 5.

2. Review of MREM and MCD

In this section, we review marginalized random effects models (MREMs) for longitudinal binary data and modified Cholesky decomposition approach (Lee and Sung, 2014).

2.1. MREM

Suppose that Y_{it} denotes the response for subject i ($i = 1, \cdots, N$) at time t ($t = 1, \cdots, n_i$) and x_{it} denotes the vector of covariates corresponding to Y_{it}. We assume that the responses for same subject are conditionally independent given random effects b_{it}, and the responses for different subjects are independent. We also assume that the conditional distribution of Y_{it} belongs to the exponential family given the random effects b_{it}, taking the form

$$P(y_{it}; b_{it}) = \exp\left\{\frac{(y_{it}\theta_{it} - \psi(\theta_{it}))}{a(\phi)} + c(y_{it}, \phi)\right\}, \tag{2.1}$$

where $\psi(\cdot)$, $a(\cdot)$, and $c(\cdot)$ are known functions, and ϕ is a scale parameter. Then MREMs are given by

$$\text{Marginal mean model:} \qquad g({\mu_{it}}^M) = x_{it}^T\beta, \tag{2.2}$$

$$\text{Dependence model:} \qquad g({\mu_{it}}^c(b_{it})) = \triangle_{it} + b_{it}, \tag{2.3}$$

$$b_i \sim \text{i.i.d. } N(0, \Sigma_i), \tag{2.4}$$

where $\mu_{it}^M = E(Y_{it}; x_{it})$ and $\mu_{it}^c(b_{it}) = E(Y_{it}; b_{it}, x_{it})$, $g(\cdot)$ is the link function, β is the $r \times 1$ vector of regression coefficients, x_{it} is a $r \times 1$ vector of covariates for subject i at time t, $b_i = (b_{i1}, \cdots, b_{in_i})^T$ and Σ_i is a $n_i \times n_i$ matrix. The parameter $\triangle_{it}$ is a function of both the marginal mean parameter, β, and the random effects variance, $var(b_{it}) = \sigma_{itt}$. From (2.2)−(2.4), $\triangle_{it}$ can be obtained as the solution of the equation

$$\mu_{it}^M = \int \mu_{it}^c(b_{it}) f(b_{it}) db_{it}, \tag{2.5}$$

where $f(b_{it})$ is normal distribution with mean 0 and variance σ_{itt}. Given (β, σ_{itt}), this equation can be solved for $\triangle_{it}$ using numerical integration and Newton-Raphson algorithm.

The MREMs have several desirable features. First, interpretation of the parameters in the marginal mean model does not depend on structure of the dependence model because the marginal and dependence models are separated. Second, estimation of covariate effects is robust to misspecification of the random effects distribution (Heagerty, 1999; Heagerty and Kurland, 2001). Third, the random effect b_i catch both the correlation between the response.

2.2. Modified Cholesky Decomposition

The RECM Σ_i is high-dimensional and it should be positive definite. To solve these problems, we use the modified Cholesky decomposition. The random effect b_i is assumed that b_{it} is regressed on its predecessors $b_{i1}, \cdots, b_{i,t-1}$. More precisely, we have

$$b_{i1} = e_{i1}, \tag{2.6}$$

$$b_{it} = \sum_{j=1}^{t-1} \phi_{i,tj} b_{ij} + e_{it}, \text{ for t=2}, \cdots, n_i. \tag{2.7}$$

Now we assume that $e_i = (e_{i1}, \cdots, e_{in_i})^T \sim N(0, D_i)$ where $D_i = diag(\sigma_{i1}^2, \cdots, \sigma_{in_i}^2)$. Then, for $t = 1, \cdots, n_i$, we have the following matrix form

$$T_i b_i = e_i, \tag{2.8}$$

where T_i is an unique unit lower triangular matrix with $1's$ on its diagonal and $-\phi_{i,tj}$ at its $(t,j)th$ position for $j < t$. Then, we have

$$T_i \Sigma_i {T_i}^T = D_i. \tag{2.9}$$

Here, ϕ_i and σ_{it}^2 are referred as generalized autoregressive parameters (GARPs) and innovation variances (IVs), respectively. For Σ_i to be positive definite the IV must be positive. The parameters GARPs and IVs can be modeled using covariate vectors $\omega_{i,tj}$ and $h_{i,t}$ by setting

$$\phi_{i,tj} = \omega_{i,tj}^T \gamma, \quad \log(\sigma_{it}^2) = h_{it}^T \lambda, \tag{2.10}$$

where γ is $a \times 1$ vectors of unknown dependence parameters, λ is $b \times 1$ vectors of unknown variance parameters, and $\omega_{i,tj}$ and $h_{i,t}$ are design vectors of covariates. The design vectors make the GARP/IV parameters depend on subject-specific covariates (Pourahmadi, 2000; Pourahmadi and Daniels, 2002; Lee *et al.*, 2012). Since all positive σ_{it}^2 guarantee the positive definiteness of Σ_i, the loglinear model in (2.10) is used. Consequently the matrix can be heterogeneous in the subject-specific covariates.

3. Partial Autocorrelation Approach

Since the modified Cholesky decomposition approach is used for modeling of the covariance matrix, we cannot use it for the correlation matrix. Instead, the partial autocorrelation approach is proposed for modeling of serial autocorrelation and this approach re-parameterizes the correlation matrix using partial autocorrelations (Daniels and Pourahmadi, 2009; Lee *et al.*, 2013).

3.1. Partial Autocorrelation

The RECM Σ_i in (2.4) can be written as the product of diagonal matrix of standard deviations D_i and correlation between responses over time R_i, that is, $\Sigma_i = D_i R_i D_i$ where $D_i = \text{diag}\{\sigma_{i1}, \cdots, \sigma_{in_1}\}$ and $R_i = (\rho_{ijk})$ is correlation matrix which has diagonal entries 1 and must be positive definite with elements taking values between -1 and 1. Therefore, the MCD can hardly satisfy the requirements.

Now we present the partial autocorrelation approach. R_i consists of its submatrices that is given by

$$R[j : j+k] = \begin{pmatrix} 1 & r_1^T(j,k) & \rho_{j,j+k} \\ r_1(j,k) & R_2(j,k) & r_3(j,k) \\ \rho_{j,j+k} & r_3^T(j,k) & 1 \end{pmatrix}, \tag{3.1}$$

where $r_1^T(j,k) = (\rho_{i,j,j+1}, \cdots, \rho_{i,j,j+k-1})$, $r_3^T(j,kl) = (\rho_{i,j+k,j+1}, \cdots, \rho_{i,j+k,j+k-1})$ and $R_2(j,k)$ is the correlation matrix corresponding to components $(j+1, \cdots, j+k-1)$ of

R_i. Now we consider a loglinear model for the standard deviation σ_{it} which is presented in Lee *et al.* (2013):

$$\log \sigma_{it} = z_{it}^T \alpha \tag{3.2}$$

where α is a $m \times 1$ coefficient vector of z_{it} and z_{it} is a subset of x_{it} allowing heterogeneity to depend on subject-level covariates such as treatment or gender.

The estimation of correlation matrix R_i is not simple because of its positive-definiteness (Daniels and Pourahmadi, 2009). To satisfy positive-definiteness of the correlation, a relatively simple structure is assumed such as AR(1) (Heagerty, 1999; Lee and Daniels, 2008; Lee *et al.*, 2009). However, it is often too strong an assumption and the covariance matrix may differ by measured covariates in many situations. Either assuming a simple structure for Σ_i or ignoring this heterogeneity can result in severe biases of estimates for the fixed effects (Lee *et al.*, 2012; Heagerty and Kueland, 2001).

To allow more flexible structures, we re-parameterize the correlation matrix R_i in terms of lag-1 $\rho_{i,j,j+1}$ and the partial autocorrelations (Joe, 2006; Daniels and Pourahmadi, 2009), $\pi_{i,j,k} = \text{cor}(Y_{ij}, Y_{ik} \mid Y_{il}, j < l < k)$, the partial correlations between Y_{ij} and Y_{ik} adjusted for the intervening variables, $\{j+1, \cdots, k-1\}$, for $j-k \geq 2$. We let Π_i the partial autocorrelation matrix with $\pi_{i,j,k}$. Then we have the following relationship between the correlation and the partial autocorrelations (Lee *et al.*, 2013),

$$\pi_{i,j,j+k} = \frac{\rho_{i,j,j+k} - r_1^T(j,k) R_2(j,k)^{-1} r_3(j,k)}{\left[1 - r_1^T(j,k) R_2(j,k)^{-1} r_1(j,k)\right]^{1/2} \left[1 - r_3^T(j,k) R_2(j,k)^{-1} r_3(j,k)\right]^{1/2}}. \tag{3.3}$$

From (3.3), we can find

$$\rho_{i,j,j+l} = r_1^T(j,l) R_2(j,l)^{-1} r_3(j,l) + D_{jl} \pi_{i,j,j+l}, \tag{3.4}$$

where D_{jl} is the denominator of the expression in (3.3).

We note that the partial autocorrelation matrix Π_i is not required to be positive-definite and its entries are free to vary in the interval (-1, 1). We also note that the equation (3.3) clearly establishes a one-to-one correspondence between the matrices R_i and Π_i. This allows swapping the constrained matrix R_i by simpler matrix Π_i.

Now we consider a transformation of the off-diagonal elements of the matrix Π_i to the entire real line using Fisher's Z-transformation which is given by

$$\frac{1}{2} \log \left(\frac{1 + \pi_{i,j,l}}{1 - \pi_{i,j,l}} \right) = \omega_{i,jk}^T \gamma, \tag{3.5}$$

where γ is $(q \times 1)$-dimensional unknown parameter and $\omega_{i,jk}$ is a unit-specific covariate vector. The choice of $\omega_{i,jk}$ makes general structure of the correlation matrix R_i. The models in (2.10) and (3.5) reduce the number of parameters of the covariance matrix, Σ_i. For example, when n_i is 4, the number of elements in Σ_i is 10. If we specify $\omega_{j,jk} = (I(|k-j| = 1), I(|k-j| = 1) \times gender)$ corresponding to an AR(1) structure depending on the covariate, gender, the number of parameters reduce from 10 to 3.

3.2. Maximum Likelihood Estimation

In this subsection, we describe the maximum likelihood estimator of $\theta = (\beta, \alpha, \gamma)$. For description, we consider longitudinal binary data. Then the link function $g(\cdot)$ is the logit function. The parameter Δ_{it} in (2.3) are functions of β in (2.2) and (α, γ) in (2.4). Given these parameters, Δ_{it}'s are calculated using a Newton-Raphson algorithm. Detailed calculations of Δ_{it} are given in Appendix C.

The MREMs likelihood function is

$$L(\theta; y) = \prod_{i=1}^{N} \int \left\{ \prod_{t=1}^{n_i} P(y_{it}; b_{it}) \right\} f(b_i) db_i. \tag{3.6}$$

To evaluate integration in (3.6), Monte Carlo integration is used. Then the log-likelihood is

$$\log L(\theta; y) = \sum_{i=1}^{N} \log \int L(\theta, b_i; y_i) f(b_i) db_i, \tag{3.7}$$

where $f(b_i)$ is a multivariate normal density with mean vector 0 and covariance matrix Σ_i and

$$L(\theta, b_i; y_i) = \exp \left[\sum_{i=1}^{n_i} \{ y_{it}(\Delta_{it} + b_{it}) + \log(1 - P_{it}^c(b_{it})) \} \right], \tag{3.8}$$

where $P_{it}^c(b_{it})) = E(Y_{it}; b_{it})$.

Detailed calculations of $\log L(\theta; y)$ are given in appendix A. Maximizing the log-likelihood with respect to θ yields the likelihood equation

$$\sum_{i=1}^{N} \frac{\partial \log L(\theta; y_i)}{\partial \theta} = \sum_{i=1}^{N} L^{-1}(\theta; y_i) \int \frac{\partial L(\theta, b_i; y_i)}{\partial \theta} f(b_i) db_i = 0. \tag{3.9}$$

The matrix of second derivatives of the observed data log-likelihood is not closed form. Therefore we use quasi-Newton methods to solve the likelihood equations, using

$$\theta^{(m+1)} = \theta^{(m)} + \left[H(\theta^{(m)}; y) \right]^{-1} \frac{\partial \log L}{\partial \theta^{(m)}}, \tag{3.10}$$

where $H(\theta)$, an empirical and consistent estimator of the information matrix at step m, is given by

$$H(\theta; y_i) = \sum_{i=1}^{N} \frac{\partial L(\theta; y_i)}{\partial \theta} \frac{\partial L(\theta; y_i)}{\partial \theta^T}. \tag{3.11}$$

When this algorithm converges, sample covariance matrix of the parameter estimates is obtained as the inverse matrix of $H(\theta; y_i)$. For the detailed calculations of maximizing the log-likelihood for the partial autocorrelation approach, see the Appendix B.

4. Real Data Analysis

In this section, we describe the real data and analyze them using our proposed model.

4.1. Description of Data

Korean Genomic Epidemiology Study (KoGES) was a cohort study to check the development of metabolic syndrome for middle-aged Korean adults aged 36-69 years. The study started in 2001 and participants were examined a total of 4 times every two years. Metabolic syndrome causes high risk for developing type 2 diabetes (Wannamethee *et al.*, 2006), coronary heart disease, and other diseases related to plaque buildups in artery walls (e.g. stroke and peripheral vascular disease).

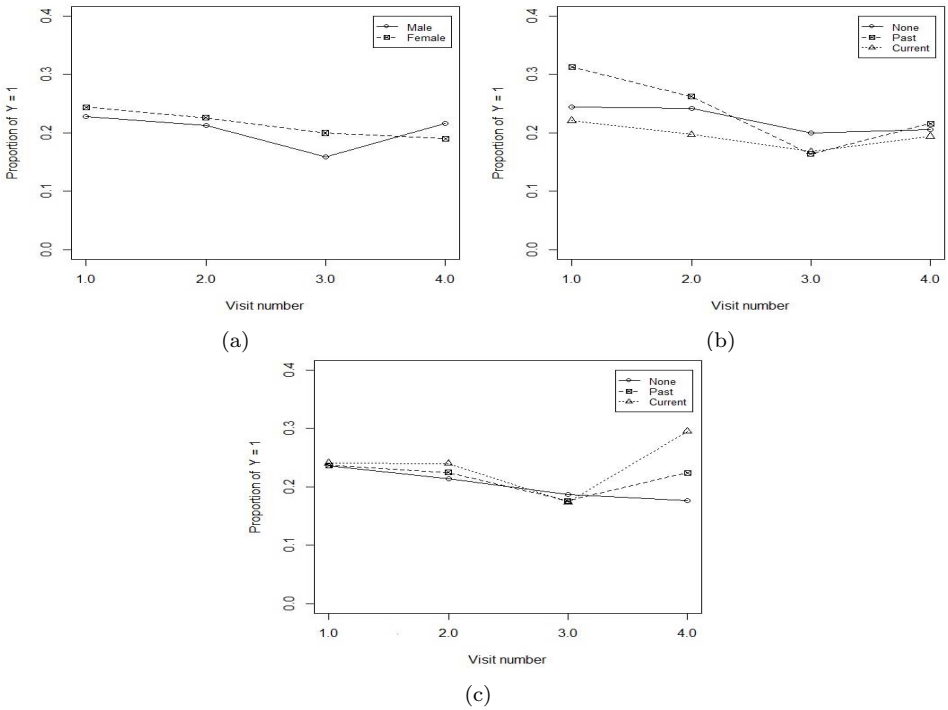

Figure 4.1 (a), (b) and (c) present sex, drinking type, and smoking type

Figure 4.1 shows the marginal prevalence of metabolic syndrome over four visits. Overall trend of the prevalence of metabolic syndrome gradually reduced over time. Then it increased at visit 4. The marginal prevalence of metabolic syndrome for females was higher than that for males and higher for the past drinking group than non-drinking group. There was no trend for smoking groups.

The response variable is to be 1 if a participant has the metabolic syndrome and 0 otherwise. Covariates included sex (female is baseline category), age (log(age/10)), alcohol drinking types (Drink 1 is the indicator for drinking in the past, Drink 2 is the indicator for drinking currently, non-drinking group is baseline category), and smoking types (Smoke 1 is the indicator for smoking in the past, Smoke 2 is the indicator for smoking currently, non-smoking group is baseline category).

There was missingness in the data. About 38% of participants dropped out before completion of the study. However, we assume the missingness is ignorable in the analysis.

4.2. Model Fit

We fit eight MREMs proposed in Sections 2 and 3. All the models are specified by $w_{i,tj}$ and $h_{i,t}$ (or z_{it}) for the construction of the RECM (Table 4.1). PA 1-4 are MREMs with the RECM of the AR(1) structure using the partial autocorrelation approach. PA 1 has a homogeneous RECM. PA 2-4 allow the RECM depending on sex, drinking type, and smoking type, respectively. MC 1-4 are MREMs with the AR(1) using the modified Cholesky decomposition approach. MC 1 has a homogeneous RECM and MC 2-4 have random effects covariance matrices depending on sex, drinking type, and smoking type, respectively.

Table 4.1 Table of eight models based on $w_{i,tj}$ and $h_{i,t}$ (z_{it})

MC 1	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1)$
MC 2	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1, \text{Sex}_i)$
MC 3	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1, \text{Drink1}_{it}, \text{Drink2}_{it})$
MC 4	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1, \text{Smoke1}_{it}, \text{Smoke2}_{it})$
PA 1	$w_{i,t,t-1} = (1)$	$z_{it} = (1)$
PA 2	$w_{i,t,t-1} = (1)$	$z_{it} = (1, \text{Sex}_i)$
PA 3	$w_{i,t,t-1} = (1)$	$z_{it} = (1, \text{Drink1}_{it}, \text{Drink2}_{it})$
PA 4	$w_{i,t,t-1} = (1)$	$z_{it} = (1, \text{Smoke1}_{it}, \text{Smoke2}_{it})$

Maximized loglikelihoods and AICs for PA 1-4 and MC 1-4 are given in Table 4.2. The likelihood ratio test for the comparison of MC 1 versus 2, MC 1 versus 3, and MC 1 versus 4 indicated that MC 1 fit significantly better than MC 2 ($\triangle D_{12} = 2 \times (3162.451 - 3162.356) = 0.190$, $p-$value$= 0.663$ on 1 d.f.), MC 3 ($p-$value$= 0.763$), and MC 4 ($p-$value$= 0.829$). Similarly we conducted the likelihood ratio test for the comparison of PA 1 versus 2, PA 1 versus 3, and PA 1 versus 4, respectively. Then we reached that PA 4 fit significantly better than PA 1-3. Using one of penalized model selection criteria, AIC, PA 4 and MC 1 were compared and we concluded that PA 4 provided a better fit than MC 1 (6342.902 and 6313.504 for MC 1 and PA 4, respectively). These comparisons indicated that PA 4 fit best among the eight models.

Table 4.2 Maximized loglikelihood values and AICs of models

Model	MC 1	MC 2	MC 3	MC 4	PA 1	PA 2	PA 3	PA 4
Loglik.	-3162.451	-3162.356	-3162.180	-3162.264	-3149.005	-3148.889	-3146.620	-3145.572
AIC	6342.902	6344.712	6346.36	6346.528	6316.01	6317.778	6315.240	6313.504

The maximum likelihood estimates of parameters for MC 1, 2 and PA 1, 4 are presented in Table 4.3. In the four models, the coefficients of covariates, gender (Sex), age (log(Age/10)), smoking type (Smoke1, Smoke2), were statistically significant (5% significance level). This indicates that the estimated marginal probability of metabolic syndrome was higher for females than for males and was higher in past-smoking group and in current smoking group than in nonsmoking group. The estimated probability of metabolic syndrome increases as age increases.

Since PA 4 was better than the other models in the partial autocorrelation and the modified Cholesky decomposition approaches, respectively, this indicates the random effects covariance matrices with a AR(1) structure depending on smoking type.

Table 4.3 Maximum likelihood estimates for MREMs
(Parameter estimates with standard errors in the parentheses)

	MC 1	MC 2	PA 1	PA 4
Marginal mean parameters: β				
Intercept	−5.055* (0.409)	−5.053* (0.410)	−5.128* (0.399)	−5.086* (0.403)
Sex (male vs female)	−0.302* (0.113)	−0.303* (0.113)	−0.311* (0.112)	−0.305* (0.112)
log(Age/10)	2.323* (0.243)	2.322* (0.244)	2.352* (0.237)	2.328* (0.239)
Drink1 (past)	−0.016 (0.118)	−0.016 (0.118)	−0.013 (0.125)	−0.016 (0.124)
Drink2 (current)	−0.120 (0.077)	−0.120 (0.077)	−0.128 (0.078)	−0.128 (0.078)
Smoke1 (past)	0.300* (0.121)	0.301* (0.121)	0.307* (0.122)	0.303* (0.123)
Smoke2 (current)	0.497* (0.119)	0.497* (0.119)	0.533* (0.120)	0.531* (0.120)
λ for MC 1,2; α for PA 1,2				
Intercept	2.537* (0.404)	2.546* (0.435)	1.840* (0.352)	1.832* (0.375)
Sex		−0.025 (0.495)		
Smoke1 (past)				0.097 (0.555)
Smoke2 (current)				−0.337 (0.388)
γ				
Intercept	0.981* (0.053)	0.981* (0.054)	0.921* (0.067)	0.947* (0.071)

* indicates significance at the 5% level of significance.

5. Conclusion

In this paper, we proposed general RECMs for marginalized random effects models using the partial autocorrelation approach. This approach compared with the covariance matrix using modified Cholesky decomposition through the real data analysis. Both approaches allow heterogeneous covariance matrix depending on subject-specific covariates and avoid the concern that the estimated covariance is not positive definite.

For parameter estimations, the likelihood approach using a quasi-Newton algorithm was used for estimation of our proposed models. To evaluate the marginalized likelihood, Monte Carlo integration is used.

In real data analysis, we fitted eight models and compared them. As a result, PA 4 fit best among the eight models and it was found that subject's gender, age, smoking status were statistically significant on metabolic syndrome. In addition, metabolic syndrome is more prevalent in females and as age increased. Smoking in the past and the current increase the chance of metabolic syndrome.

Appendix A: Calculations of quasi-Newton for partial autocorrelation approach

Let $\theta = (\beta^T, \alpha^T, \gamma)$. Then the log likelihood is given by

$$\log L(\theta; y_i) \quad = \quad \sum_{i=1}^{N} \log \int L(\theta, b_i; y_i) f(b_i) db_i,$$

where

$$L(\theta, b_i; y_i) \quad = \quad \exp\left[\sum_{t=1}^{n_i} \{y_{it}(\Delta_{it} + b_{it}) + \log(1 - P_{it}^{c}(b_{it}))\}\right].$$

Appendix B: Detailed Calculations of the quasi-Newton Method for MREMs

For β, the derivative of the marginal log–likelihood is

$$\frac{\partial \log L(\theta;y)}{\partial \beta} = \sum_{i=1}^{N} \frac{1}{L(\theta;y_i)} \int L(\theta,b_i;y_i) \sum_{t=1}^{n_i} \{y_{it} - P_{it}^c(b_{it})\} \frac{\partial \Delta_{it}}{\partial \beta} f(b_i) db_i,$$

Similar to $\frac{\partial \log L(\theta;y)}{\partial \beta}$, $\frac{\partial \log L(\theta;y)}{\partial \alpha_l}$ and $\frac{\partial \log L(\theta;y)}{\partial \gamma_k}$ are given as follows.

$$\begin{aligned}
\frac{\partial \log L(\theta;y)}{\partial \alpha_l} &= \sum_{i=1}^{N} \frac{1}{L(\theta;y_i)} \int L(\theta,b_i;y_i) \Bigg[\sum_{t=1}^{n_i} (y_{it} - P_{it}^c(b_{it})) \frac{\partial \Delta_{it}}{\partial \alpha_l} \\
&\quad - \sum_{t=1}^{n_i} z_{it} + \frac{1}{2} \sum_{t=1}^{n_i} b_i^T \left\{ D^{-1} \frac{\partial D_i}{\partial \alpha_l} D_i^{-1} R_i^{-1} D_i^{-1} + D_i^{-1} R_i^{-1} D_i^{-1} \frac{\partial D_i}{\partial \alpha_l} D_i^{-1} \right\} b_i \Bigg] f(b_i) db_i, \\
\frac{\partial \log L(\theta;y)}{\partial \gamma_k} &= \sum_{i=1}^{N} \frac{1}{L(\theta;y_i)} \int L(\theta,b_i;y_i) \Bigg[-\frac{1}{2} \sum_{1 \le t < k \le n_i} Tr \left(R_i^{-1} \frac{\partial R_i}{\partial \rho_{i,tk}} \right) \frac{\partial \rho_{i,tk}}{\partial \pi_{i,tk}} \frac{\partial \pi_{i,tk}}{\partial \gamma_k} \\
&\quad + \frac{1}{2} \sum_{1 \le t < k \le n_i} b_i^T D_i^{-1} R_i^{-1} \frac{\partial R_i}{\partial \rho_{i,tk}} R_i^{-1} D_i^{-1} b_i \frac{\partial \rho_{i,tk}}{\partial \pi_{i,tk}} \frac{\partial \pi_{i,tk}}{\partial \gamma_k} \Bigg] f(b_i) db_i,
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial D_i^{-1}}{\partial \alpha_l} &= -diag \left\{ \frac{z_{i1l}}{\sigma_{i1}}, \cdots, \frac{z_{in_i l}}{\sigma_{in_i}} \right\}, \\
\frac{\partial \pi_{i,tk}}{\partial \gamma_k} &= \omega_{i,tk} (1 - \pi_{i,tk})^2.
\end{aligned}$$

Here, derivatives of Δ_{it} are calculated using the following relationship,

$$P_{it}^M = \int P_{it}^c(b_{it}) f(b_i) db_i.$$

Now taking the derivative with respect to β, we get

$$\frac{\partial \Delta_{it}}{\partial \beta} = \frac{P_{it}^M (1 - P_{it}^M) x_{it}}{\int P_{it}^c(b_{it})(1 - P_{it}^c(b_{it})) f(b_i) db_i}.$$

Similarly, taking the derivative with respect to α, we get

$$\frac{\partial \Delta_{it}}{\partial \alpha_l} = -\frac{\int P_{it}^c(b_{it})(1 - P_{it}^c(b_{it})) b_i \sigma_{it} z_{it} f(b_i) db_i}{\int P_{it}^c(b_{it})(1 - P_{it}^c(b_{it})) f(b_i) db_i}.$$

Appendix C: Calculation of Δ_{it}

The intercepts Δ_{it} are a function of β, α and λ and must be obtained within the Newton-Raphson algorithm. Let $h(\Delta_{it}) = \int P_{it}^c(b_{it})^{y_{it}} f(b_i) db_i - P_{it}^M$. Then Δ_{it} can be obtained by solving this Newton-Raphson algorithm:

$$\Delta_{it}^{(n+1)} = \Delta_{it}^{(n)} - \left(\frac{\partial h(\Delta_{it})}{\partial \Delta_{it}}\right)^{-1} h(\Delta_{it}),$$

where

$$\frac{\partial h(\Delta_{it})}{\partial \Delta_{it}} = \int P_{it}^{c}(b_{it})(1 - P_{it}^{c}(b_{it}))f(b_{it})db_{it}.$$

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., Wiley and Sons, New York.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 125-134.

Choi, N. and Huh, J. (2014). A longitudinal study for child aggression with Korea welfare panel study data. *Journal of the Korean Data & Information Science Society*, **25**, 1439-1447.

Daniels, M. J. and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, **89**, 553-566.

Daniels, M. J. and Pourahmadi, M. (2009). Modelling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis*, **100**, 2352-2363.

Daniels, M. J. and Zhao, Y. D. (2003). Modeling repeated count data subject to informative dropout. *Statistics in Medicine*, **22**, 1631-1647.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary response. *Biometrika*, **80**, 141-151.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688-698.

Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, **58**, 342-351.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, **88**, 973-985.

Jeon, J. and Lee, K. (2014). Review and discussion of marginalized random effects models. *Journal of the Korean Data & Information Science Society*, **25**, 1263-1272.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, **97**, 2177-2189.

Lee, K. and Daniels, M. J. (2008). Marginalized models for longitudinal ordinal data with application to quality of life studies. *Statistics in Medicine*, **27**, 4359-4380.

Lee, K., Joo, Y., Yoo, J. K. and Lee, J. (2009). Marginalized random effects models for multivariate longitudinal binary data. *Statistics in Medicine*, **28**, 1287-1300.

Lee, K. and Mercante, D. (2010). Longitudinal nominal data analysis using marginalized models. *Computational Statistics and Data Analysis*, **54**, 208-218.

Lee, K., Kang, S., Liu, X. and Seo, D. (2011). Likelihood-based approach for analysis of longitudinal nominal data using marginalized random effects models. *Journal of Applied Statistics*, **38**, 1577-1590.

Lee, K., Lee, J., Hagan, J and Yoo, J. K. (2012). Modelling the random effects covariance matrix for generalized linear mixed models. *Computational Statistics and Data Analysis*, **56**, 1545-1551.

Lee, K., Daniels, M. J. and Joo, Y. (2013). Flexible marginalized models for bivariate longitudinal ordinal data. *Biostatistics*, **14**, 462-476.

Pan, J. and MacKenzie, G. (2003). On modelling mean-covariance structure in longitudinal studies. *Biometrika*, **90**, 239-244.

Pan, J. and MacKenzie, G. (2007). Modelling conditional covariance in the linear mixed model. *Statistical Modelling*, **7**, 49-71.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, **86**, 677-690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425-435.

Wang, Y. and Daniels, M. J. (2013). Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances. *Journal of Multivariate Analysis*, **116**, 130-140.

Wannamethee, S. G., Shaper, A. G., Lennon, L., Morris, R. W. (2006). Metabolic syndrome vs Framingham risk score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. *Journal of the American Medical Association*, **295** 819-821.