Communications for Statistical Applications and Methods
2014, Vol. 21, No. 2, 169–181

DOI: http://dx.doi.org/10.5351/CSAM.2014.21.2.169
ISSN 2287-7843

Autoregressive Cholesky Factor Modeling for Marginalized Random Effects Models

Keunbaik Lee[1,a], Sunah Sung[b]

[a]Department of Statistics, Sungkyunkwan University, Korea; [b]GS HomeShopping Inc., Korea

Abstract

Marginalized random effects models (MREM) are commonly used to analyze longitudinal categorical data when the population-averaged effects is of interest. In these models, random effects are used to explain both subject and time variations. The estimation of the random effects covariance matrix is not simple in MREM because of the high dimension and the positive definiteness. A relatively simple structure for the correlation is assumed such as a homogeneous AR(1) structure; however, it is too strong of an assumption. In consequence, the estimates of the fixed effects can be biased. To avoid this problem, we introduce one approach to explain a heterogenous random effects covariance matrix using a modified Cholesky decomposition. The approach results in parameters that can be easily modeled without concern that the resulting estimator will not be positive definite. The interpretation of the parameters is sensible. We analyze metabolic syndrome data from a Korean Genomic Epidemiology Study using this method.

Keywords: Population-averaged effect, heterogeneity, Quasi-Monte Carlo, autoregressive model, positive definite.

1. Introduction

In longitudinal data analysis, likelihood-based approaches have been used frequently such as generalized linear mixed models (Breslow and Clayton, 1993) and marginalized random effects models (Heagerty, 1999). In these models random effects are used to explain the serial correlation of the repeated measurements from the same subject and the random effects covariance matrix must be taken into account for proper inference on covariate effects (Fitzmaurice and Laird, 1993). The random effects covariance matrix is assumed that it is constant over subjects and is restricted because of its high dimensionality and the positive definite constraint. However, in many situations, these assumptions are too strong and result in biased estimates of the fixed effects (Heagerty and Kurland, 2001).

Recently the parameters of the random effects covariance matrix were proposed to depend on subject-specific covariates using two methods: the partial autocorrelation approach and the modified Cholesky decomposition. The partial autocorrelation approach uses a more flexible class of models for serial autocorrelation by re-parameterizing the correlation matrix using partial autocorrelations (Daniels and Pourahmadi, 2009; Lee *et al.*, 2013). The partial autocorrelation matrix is not required to be positive-definite and the correlation matrix corresponding to it is positive-definite. Wang and Daniels (2013) discussed priors for parameters of the partial autocorrelations in Bayesian versions of correlation models. This approach was extended to analyze bivariate longitudinal ordinal data in

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (2012-004002).

[1] Corresponding author: Associate Professor, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 110-745, Korea. E-mail: keunbaik@skku.edu

Published online 31 March 2014 / journal homepage: http://csam.or.kr

Lee *et al.* (2013). The modified Cholesky decomposition approach uses a set of dependence parameters, generalized autoregressive parameters (GARP) and a set of variance parameters, the innovation variances (IV). The positive definiteness restriction of the covariance matrix is that the IV need to be positive (Pourahmadi, 1999, 2000). Daniels and Pourahmadi (2002) proposed Bayesian priors for the GARP and IV in linear mixed models. Pourahmadi and Daniels (2002) developed dynamic conditionally mixed models by decomposing the within-subject covariance matrix using a modified Cholesky decomposition (Pourahmadi, 1999, 2000). Similar modeling for covariance matrices was proposed in Daniels and Zhao (2003). Pan and MacKenzie (2003, 2006) generalized Pourahmadi's (1999, 2000) method to deal with unbalanced longitudinal data and to address joint mean-covariance estimation for linear mixed models. Lee *et al.* (2012) extended the modified Cholesky decomposition approach to generalized linear mixed models. In this paper, we extend modified Cholesky decomposition approach for random effects covariance matrix in marginalized random effects models.

Marginalized models are commonly used for analyzing longitudinal categorical data when the population-averaged effects is of interest. Marginalized models are likelihood-based models (Heagerty, 1999, 2002; Lee and Daniels, 2007, 2008; Lee *et al.*, 2009; Lee and Mercante, 2010; Lee *et al.*, 2011). And the correlation of repeated measurements in these models is modeled via random effects (marginalized random effects models; MREMs) or a Markov correlation structure (marginalized transition models; MTM) while the population averaged response is directly modeled as a function of covariates, which induces restrictions on the correlation model. In this paper, we consider a marginalized random effects model with general random effects covariance matrix using the modified Cholesky decomposition.

The paper is organized as follows. In Section 2, we propose marginalized random effects models with autoregressive structure of random effects covariance matrix. using a modified Cholesky decomposition. In Section 3, we analyze data from the longitudinal study on metabolic syndrome. In Section 4, we conduct a simulation study to present a marginalized random effect models (MREMs) with using modified Cholesky decomposition. Finally, conclusions and extensions are provided in Section 5.

2. Marginalized Modeling for Longitudinal Categorical Data

In this section, we propose marginalized random effects models (MREMs) with autoregressive (AR) structure of random effects covariance matrix. The AR structure of the covariance matrix is explained using modified Cholesky decomposition.

2.1. Marginalized random effects models

We first explain marginalized random effects models for longitudinal categorical data. Let Y_{it} be the response for subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$). We assume that each Y_{it} is conditionally independent given random effects b_i, the responses for different subjects are independent, and Y_{it} has a conditional distribution in the exponential family given the random effects b_i, taking the form

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

where $a(\cdot)$, $\psi(\cdot)$, and $c(\cdot)$ are known functions, and ϕ is a scale parameter. We model a transformation of the mean, which would be some function of θ_{it}, as a linear model in both the fixed and random factors. A transformation of the mean would be some function of θ_{it}, as a linear model in both the fixed and random factors.

Let $\mu_{it}^M = E(Y_{it}; x_{it})$ and $\mu_{it}^c(b_{it}) = E(Y_{it}; x_{it}, b_{it})$ be the marginal mean given covariates and the conditional mean given random effects, respectively. Then the proposed model is given by

$$\text{Marginal mean model: } g\left(\mu_{it}^M\right) = x_{it}^T\beta, \tag{2.1}$$

$$\text{Dependence model: } g\left(\mu_{it}^c(b_{it})\right) = \Delta_{it} + b_{it}, \tag{2.2}$$

$$b_i \sim N(\mathbf{0}, \Sigma_i), \tag{2.3}$$

where β is the $r \times 1$ vector of marginal mean parameters, x_{it} is a $r \times 1$ vector of covariates for subject i at time t, $b_i = (b_{i1}, \ldots, b_{in_i})^T$, $\mathbf{0}$ is a vector of zeros with length n_i, and Σ_i is a $n_i \times n_i$ matrix. Let σ_{it}^2 be the t^{th} diagonal element of Σ_i. The parameters Δ_{it} in (2.2) are functions of the marginal mean parameter, β, in (2.1) and random effects variance, σ_{it}^2 in (2.3) such that

$$\mu_{it}^M = \int \mu_{it}^c(b_{it}) f(b_{it}) db_{it}, \tag{2.4}$$

where $f(b_{it})$ is the normal probability density with mean 0 and variance σ_{it}^2. Given β, and σ_{it}^2, we calculate Δ_{it} using a Newton-Raphson algorithm from (2.4). Details regarding the Newton-Raphson algorithm to find each Δ_{it} is given in Heagerty (1999).

This model has several desirable features. First, since the marginal distribution of the observed data are reasonably preserved, interpretation of the marginal mean parameters does not depend on specification of the dependence model. Second, the parameter estimation of marginal mean models is less susceptible to bias resulting from random effects model misspecification (Heagerty, 1999; Heagerty and Kurland, 2001). This is an advantage of marginal models over generalized linear mixed models. Third, the random effects b_i in (2.2) capture both the correlation between the responses. In generalized linear mixed models (GLMMs), a single scalar random intercept b_i shared by all n_i components within a subject, and does not allow a broader class of a vector of correlated random effects within each subject.

2.2. Random effects covariance matrix

The random effects covariance matrix, Σ_i in (2.3) is high dimensional and should be positive definite. Therefore, its structure is assumed to be constant over subjects and to be restricted such as AR(1) structure. However, these assumptions are too strong and can result in biased estimates of the fixed effects. In this subsection, we propose modified Cholesky decomposition approach of random effects covariance matrix for MREMs to solve the problems.

2.2.1. Modified Cholesky decomposition

We now propose a modified Cholesky decomposition approach for random effects covariance matrix of the MREMs. The key idea of the modified Cholesky decomposition is that the covariance matrix Σ_i of the random vectors b_i in (2.3) can be diagonalized by a lower triangular matrix constructed from the regression coefficients when b_{it} is regressed on its predecessors $b_{i1}, \ldots, b_{it-1}$. More precisely, for $t = 2, \ldots, n_i$, we have

$$b_{it} = \sum_{j=1}^{t-1} \phi_{i,tj} b_{ij} + e_{it}, \quad T_i \Sigma_i T_i^T = D_i, \tag{2.5}$$

where T_i and D_i are unique matrices, T_i is a unit lower triangular having ones on its diagonal and $-\phi_{i,tj}$ at its $(t, j)^{th}$ position for $j < t$, and D_i is diagonal with $\sigma_{i,t}^2 = \text{var}(e_{it})$ as its diagonal entries. The ϕ is

referred as generalized autoregressive parameters (GARP) and the $\sigma^2_{i,t}$ as innovation variances (IV). The constraint on theses parameters for Σ_i to be positive definite is that the IV need to be positive.

The standard Cholesky decomposition of a positive definite matrix is of the form

$$\Sigma_i = L_i L_i^T, \tag{2.6}$$

where L_i is lower triangular with positive diagonal elements. When $L_i = T_i^{-1} D_i^{1/2}$, (2.6) is rewritten as

$$\Sigma_i = T_I^{-1} D_i T_i^{-T},$$

which is the result of the modified Cholesky decomposition of Σ_i.

In the standard Cholesky decomposition, statistical interpretation of the entries of L_i in (2.6) is difficult in its present form (Pinheiro and Bates, 1996). However, in the modified Cholesky decomposition, the parameters, GARP and IV have meaningful statistical interpretation. The parameters, GARP and IV can be modeled using time and/or subject-specific covariate vectors $w_{i,tj}$ and $h_{i,t}$ by setting

$$\phi_{i,tj} = w_{i,tj}^T \gamma, \qquad \log(\sigma^2_{i,t}) = h_{i,t}^T \lambda, \tag{2.7}$$

where γ and λ are $a\times 1$ and $b\times 1$ vectors of unknown dependence and variance parameters, respectively. Note that design vectors $w_{i,tj}$ and $h_{i,t}$ are used to model the GARP/IV parameters as functions of subject-specific covariates. These models were commonly used for random effects covariance matrix in the linear mixed models to allow the covariance matrix to vary across subjects (Pourahmadi, 2000; Pourahmadi and Daniels, 2002; Daniels and Zhao, 2003).

The modified Cholesky decomposition decomposes the parameters of Σ_i to the GARP and log(IV)'s. The GARP/IV parameterization has several advantages. First, since the GARP and log(IV) are unconstrained, we can model the covariance matrix in terms of covariates. Second, the parameters have a sensible interpretation because of linear combination of covariates in (2.7). Finally, the GARP/IV parameterization in (2.7) provides parameters that can be easily modeling.

2.3. Maximum likelihood estimation

To derive the likelihood function for the (2.1)–(2.3) we use the simple case of longitudinal binary data. Then the link function $g(\cdot)$ is the logit. Let $\theta = (\beta, \gamma, \lambda)$. Then the parameters Δ_{it} in (2.2) are functions of β in (2.1) and (γ, λ) in (2.3). Given these parameters, we calculate Δ_{it} using a Newton-Rapson algorithm in Heagerty (1999). Detail calculations are given in Appendix . The likelihood function is the integral over random effects of a product of Bernoulli distributions.

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

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

Then the log-likelihood is given by

$$\log L(\theta; y) = \sum_{i=1}^{N} \log \int L(\theta, b_i; y_i) f(b_i)\, db_i, \tag{2.8}$$

where

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

Maximizing the log-likelihood with respect to θ yields the likelihood equations

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

Since the analytic forms of second derivatives of the observed data log-likelihood in (2.8) are not closed forms, we use the quasi-Newton methods to solve the likelihood equations. The $(c + 1)$th iteration $\theta^{(c+1)}$ is updated using

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

where $H(\theta; y, r)$ is computed by the second derivatives of the log marginal likelihood. However, it is not easy to compute the second derivatives of log marginal likelihood because of the integrals of random effects. Instead, we use the product of the first derivatives of log marginal likelihood. Therefore, we use an empirical and consistent estimator of the information matrix at step (c) and is given as

$$H(\theta; y, r) = \sum_{i=1}^{N} \frac{\partial L(\theta; y_i, r_i)}{\partial \theta} \frac{\partial L(\theta; y_i, r_i)}{\partial \theta^T}.$$

At convergence, the large-sample variance-covariance matrix of the parameter estimates can be obtained as the inverse of the information matrix. Details of maximizing the log-likelihood for the modified Cholesky decomposition approach is provided in the Appendix.

3. Example

3.1. Metabolic syndrome data

Our approach is motivated by data from the Korean Genomic Epidemiology Study (KoGES) (Kim *et al.*, 2006). The purpose of the KoGES is to monitor the development of metabolic syndrome for middle-aged Korean adults aged 39–69 years. Participants were examined every two years for up to eight years to monitor the development of metabolic syndrome. Metabolic syndrome is a combination of medical disorders that increase the risk of developing cardiovascular disease and diabetes. It affects one in five people, and prevalence increases with age. It is also a significant risk factor for developing type 2 diabetes, coronary heart disease, and other diseases related to plaque buildup in artery walls (*e.g.*, stroke and peripheral vascular disease).

Metabolic syndrome is defined as three or more of the following five disorders: abdominal obesity (waist circumference > 90cm in men or > 80cm in women), high blood pressure (systolic BP levels > 130mmHg or diastolic BP levels > 85mmHg), high impaired fasting glucose (IFG > 110mg/dl), high triglyceridemia (TG > 150mg/dl), and low high-density lipoprotein cholesterol (HDL-C < 40mg/dl in men or < 50mg/dl in women). It is of primary interest how demographic factors affect metabolic syndrome. The demographic factors were sex, age, alcohol consumption, and smoking.

Table 1: Summary of prevalence of Metabolic Syndrome from KoGES Data.

Prop. of $Y = 1$ (Number of obs.)		Visit				overall prop.
		1 (2310)	2 (1887)	3 (1679)	4 (1416)	
Sex	male	0.228	0.213	0.159	0.216	0.206
	female	0.244	0.226	0.200	0.190	0.219
Drink	none	0.244	0.242	0.200	0.206	0.226
	past	0.313	0.262	0.164	0.216	0.242
	current	0.221	0.197	0.168	0.194	0.197
Smoke	none	0.236	0.214	0.187	0.176	0.207
	past	0.237	0.225	0.176	0.224	0.214
	current	0.241	0.240	0.174	0.295	0.236
overall prop.		0.237	0.220	0.183	0.201	

Table 2: Table of four models based on $w_{i,tj}$ and $h_{i,t}$.

MC 1	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1)$
MC 2	$w_{i,t,t-1} = (1, \text{Sex}_i)$	$h_{i,t} = (1, \text{Sex}_i)$
GLM 1	$w_{i,t,t-1} = (1)$	$h_{i,t} = (1)$
GLM 2	$w_{i,t,t-1} = (1, \text{Sex}_i)$	$h_{i,t} = (1, \text{Sex}_i)$

Table 1 indicates that the marginal prevalence of metabolic syndrome over four visits are summarized. The marginal prevalence of metabolic syndrome was higher for females than males, higher for the current smoking group than for the past smoking group, higher for the past smoking group than for the non-smoking group, higher for past drinking group than for the non-drinking group, and higher for the non-drinking group than for the current drinking group over four visits. The baseline mean age of 945 males was 49.21 and that of 1,365 females was 49.27.

We analyzed the demographic factors in KoGES associated with metabolic syndrome. The response variable is 1 if a participant has metabolic syndrome and 0 otherwise. As predictors, we included sex (1 = male; 0 = female), age (log(age/10)), alcohol consumption types (Drink1 = 1 if drinking in the past, 0 otherwise; Drink2 = 1 if drinking currently, 0 otherwise), and smoking types (Smoke1 = 1 if smoking in the past, 0 otherwise; Smoke2 = 1 if smoking currently, 0 otherwise).

3.2. Computations

Implementing the quasi-Newton algorithm in these models was computationally intensive because the estimates required numerical integration for all subjects. We used simultaneously R software and FORTRAN. R software was used for the quasi-Newton iteration and FORTRAN was used to make subroutines(.dll Files) to conduct the numerical integrations. Each quasi-Newton step for our proposed model required approximate 40 seconds.

3.3. Model fit

We fit two marginalized random effects models proposed in Section 2. We also fit two generalized linear mixed models using the modified Cholesky decomposition in Lee *at al.* (2012). The four models used the modified Cholesky decomposition approach for Σ_i. All models had random effects covariance matrix with AR(1) structure, and the models are specified by $w_{i,tj}$ and $h_{i,t}$ in Table 2. MC 1 has a homogeneous random effects covariance matrix with AR(1) and MC 2 has random effects covariance matrixes depending on sex, respectively. AICs for MC 1–2 and GLM 1–2 are given in Table 3. Using a penalized model selection criterion (AIC) MC 1 provided a better fit than MC 2. (6322.648 for MC 1, 6326.608, 6506.654 and 6490.920 for MC 2, GLM 1–2 respectively). These comparisons indicated that MC 1 fit best among the four models.

Table 3: AICs of models

Model	MC 1	MC 2	GLM 1	GLM 2
AIC	6322.648	6326.608	6506.654	6490.920

Table 4: Maximum likelihood estimates for marginalized random effects model (Parameter estimates with standard errors in the parentheses calculated using the diagonal of inverse of Hessian matrix of $l(\theta)$ at convergence)

	MC 1	MC 2
Marginal mean parameters: β		
Intercept	−4.964* (0.412)	−4.936* (0.415)
Sex (male vs female)	−0.281* (0.107)	−0.291* (0.125)
log(Age/10)	2.303* (0.245)	2.287* (0.245)
Drink1 (past)	−0.018 (0.107)	−0.015 (0.108)
Drink2 (current)	−0.101 (0.072)	−0.100 (0.071)
Smoke1 (past)	0.293* (0.112)	0.296* (0.113)
Smoke2 (current)	0.454* (0.112)	0.459* (0.113)
λ for MC 1,2		
Intercept	2.244* (0.324)	2.202* (0.420)
Sex		0.043 (0.611)
γ		
Intercept	1.048* (0.053)	1.074* (0.075)
Sex		−0.052 (0.104)
Maximized log-likelihood	−3152.324	−3152.304
AIC	6322.648	6326.608

*Indicates significance at the 5 % level of significance.

Table 4 presents the results from two model fits (MC 1–2). In all two models, the coefficients of gender (Sex), age (log(Age/10)), smoking type (Smoke1, Smoke2) were statistically significant under 5% significance level. This suggests that the estimated marginal probability of metabolic syndrome was lower for males than females and was higher in the past-smoking group and in the current-smoking group than in the nonsmoking group. The estimated probability of Metabolic Syndrome increased as age increased. Figures 1 and 2 presents the difference of estimated marginal probabilities. Figures 1 and 2 are plots for averages of estimated marginal probabilities of metabolic syndrome according to gender and smoking groups, respectively. These figures presents the differences between male and female subjects and among smoking groups, respectively.

Since MC 1 was better than the MC 2 in the modified Cholesky decomposition approach, this indicates that the random effects covariance matrices had homogeneous AR(1) structures and the estimated value of $\Sigma_i = T_i^{-1} D_i (T_i^T)^{-1}$.

4. Simulation Study

We conducted simulations to compare structures of the random effects covariance matrix via examining biases and coverage probabilities of estimates of the marginal mean parameters. We considered 200 datasets from a MREM with two covariates, time and group (2 levels). Longitudinal binary responses were planned to take place at 6 equally spaced visit times. The marginal probability for the MREM was specified as

$$\begin{aligned} \operatorname{logit} P(Y_{it} = 1 | x_{it}) &= \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{it} + \beta_3 \text{group}_i \, \text{time}_{it} \\ \beta &= (\beta_0, \beta_1, \beta_2, \beta_3) = (0.5, -0.4, -0.1, -0.2), \end{aligned}$$

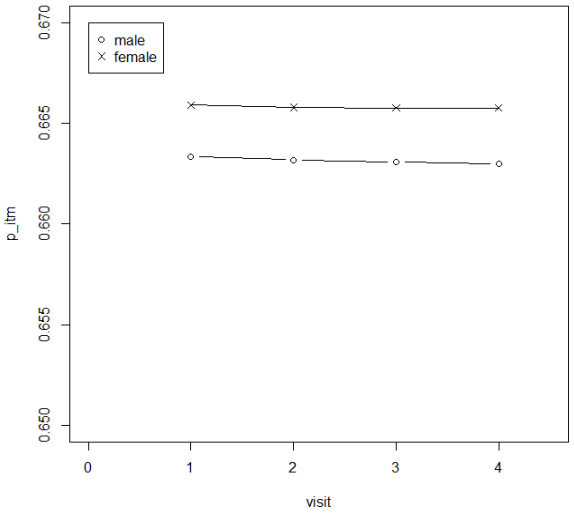

Figure 1: *Plot for averages of estimated marginal probabilities of metabolic syndrome for gender groups*

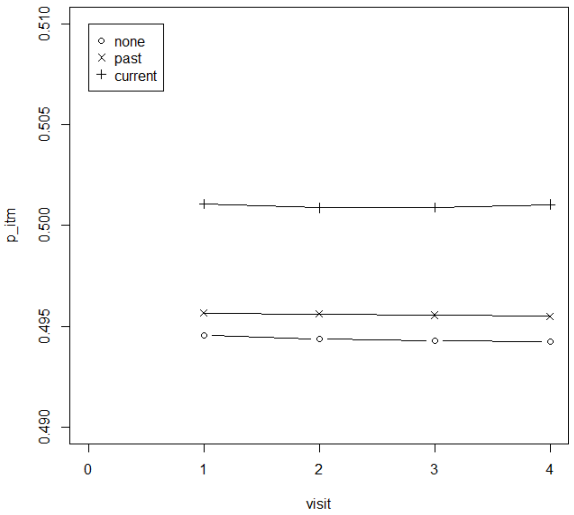

Figure 2: *Plot for averages of estimated marginal probabilities of metabolic syndrome for smoking groups*

where $\text{time}_{it} = t/10$ for $t = 0, 1, \ldots, 6$, and $\text{group}_i = 0$ or 1 with an equal sample size (100) per group. The conditional probability was specified as

$$\text{logit}\, P(Y_{it} = 1|b_{it}) = \Delta_{it} + b_{it}, \tag{4.1}$$

$$b_i \sim N(0, \Sigma_i). \tag{4.2}$$

We first considered two cases with different covariance matrices Σ_i using the modified Cholesky decomposition approach.

Case1: We generated longitudinal binary responses with sample size of 200 from marginalized ran-

Table 5: Averages and biases of fitted marginal mean parameters for Models 1 and 2 (Coverage probabilities)

	Model 1		Model 2	
	mean	bias	mean	bias
β_0 (0.5)	0.504 (0.947)	0.004	0.494 (0.955)	0.006
β_1 (−0.4)	−0.388 (0.962)	0.012	−0.398 (0.945)	0.002
β_2 (−0.1)	−0.099 (0.962)	0.001	−0.098 (0.950)	0.002
β_3 (−0.2)	−0.229 (0.923)	0.029	−0.191 (0.930)	0.009

Table 6: Averages and biases of fitted marginal mean parameters for Models 3 and 4 (Coverage probabilities)

	Model 3		Model 4	
	mean	bias	mean	bias
β_0 (0.5)	0.504 (0.955)	0.004	0.501 (0.930)	0.001
β_1 (−0.4)	−0.418 (0.950)	0.018	−0.404 (0.975)	0.004
β_2 (−0.1)	−0.099 (0.965)	0.001	−0.102 (0.940)	0.002
β_3 (−0.2)	−0.178 (0.975)	0.022	−0.199 (0.970)	0.001

dom effects model with AR(1) covariance matrix using modified Cholesky decomposition. The covariance matrix was assumed to be homogenous. That is, Σ_i had $\phi_{i,tj} = \gamma_0 I(|t-j| = 1)$ and $\log(\sigma^2_{i,t}) = \lambda_0$. The assumed values for the parameters are $\gamma_0 = 0.5$ and $\lambda_0 = 0.1$. Then we fit the true model (Model 1).

We also generated datasets from the same model with a different covariance matrix depending on group. That is, Σ_i had $\phi_{i,tj} = \gamma_0 I(|t-j| = 1) + \gamma_1 I(|t-j| = 1)\text{group}_i$ and $\log(\sigma^2_{i,t}) = \lambda_0 + \lambda_1 \text{group}_i$. The assumed values for the parameters γ and λ were $(\gamma_0, \gamma_1) = (0.5, 0.1)$ and $(\lambda_0, \lambda_1) = (0.05, 0.1)$. Then we fit model with the homogeneous AR(1) covariance matrix (Model 2).

Table 5 presents the biases and coverage probabilities of marginal mean parameter estimates. The averages of fitted marginal mean parameters for Model 1 were more close to true values than for Model 2 except β_1 and β_3. The coverage probabilities were almost same.

Case 2: We next generated datasets from marginalized random effects model with AR(2) covariance matrix using modified Cholesky decomposition. We assumed homogeneous AR(2) structure for Σ_i which was given by $\phi_{i,tj} = \gamma_0 I(|t-j| = 1) + \gamma_1 I(|t-j| = 2)$ and $\log(\sigma^2_{i,t}) = \lambda_0$. The assumed values for the parameters γ and λ were $(\gamma_0, \gamma_1) = (0.5, 0.1)$ and $\lambda_0 = 0.1$, respectively. Then we fit two marginalized random effects models with same covariance matrix (Model 3) and an AR(1) homogeneous covariance matrix (Model 4), respectively.

Table 6 presents biases and coverage probabilities of estimates of marginal mean parameters. Similar to Case 1, we had similar average fitted values for marginal mean parameters. The coverage probabilities were similar.

From above simulations, we know that even though we fit MREM with wrong covariance matrix the average biases and coverage probabilities were similar. The reason is that MREMs have a robust property of estimation of marginal mean parameters to the misspecification of dependence model under no missingness.

5. Summary and Discussion

In this paper we proposed marginalized random effects models with an autoregressive structure of covariance matrix using modified Cholesky decomposition. The covariance matrix is decomposed to the innovation variances (IVs) and generalized autoregressive parameters (GARPs). This structure allows heterogeneous covariance matrix depending on subject-level covariates and satisfies the positive-definiteness of the covariance matrix with positive IVs. Parameter estimation was based on marginalized maximum likelihood estimation using a quasi-Newton algorithm. To numerically integrate over the random effects, the quasi-Monte Carlo method was used in the likelihood equations.

In our KoGES analysis, we found that participant's gender, age, smoking status had statistically significant effects on metabolic syndrome. The proposed models capture usual population-averaged effects in marginal mean as well as within-subject serial dependence of the responses in the dependence model.

Simulation studies indicated that marginal mean parameter estimates with random effects covariance matrix using modified Cholesky decomposition approach was well applied to marginalized random effects models (MREMs). The MREMs have a robust property of estimation of marginal mean parameters to misspecification of dependence model under no missingness. However, under MAR, it is not guaranteed. We have left this case for future study.

The modified Cholesky decomposition is used for the covariance matrix. However, the correlation matrix is commonly used instead of the covariance matrix. Therefore, we will consider the marginalized random effects models with a general covariance matrix using partial autocorrelation approach. Instead of modeling of the covariance matrix, the partial autocorrelation approach uses a partial autocorrelation matrix that is not required to be positive-definite. The correlation matrix corresponding to it is positive-definite. This approach is ongoing in the future study and we will explore various methods to structure this matrix in terms of stable estimation and in terms of feasible computations.

Appendix A: Calculation of Δ_{it}

The intercepts $\triangle_{it}$ are a function of β, γ, and λ and must be obtained within the Newton-Raphson algorithm. Let $h(\triangle_{it}) = \int P_{it}^c(b_i^{(t)})f(b_i)db_i - P_{itj}^M$. Estimates of $\triangle_{it}$ can be obtained using Newton-Raphson as follows,

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

where

$$\frac{\partial h(\triangle_{it})}{\partial \triangle_{it}} = \int P_{it}^c(b_{it})\left(1 - P_{it}^c(b_{it})\right) f(b_{it})db_{it}. \tag{A.1}$$

Note that the integral in (A.1) is one-dimensional and we use QMC to evaluate this integral. Numerical evaluation of the integral in this paper is accomplished by using a 40-point Gauss-Hermite quadrature with a maximal error $< 10^{-5}$ for the real data analysis and a 40-point Gauss-Hermite quadrature with a maximal error $< 10^{-4}$ for the simulation study.

Appendix B: Calculations of quasi-Newton for modified Cholesky decomposition approach

Maximizing the log-likelihood with respect to θ yields the likelihood equations

$$\sum_{i=1}^{N} \frac{\partial \log L(\theta; y_i)}{\partial \theta} = 0,$$

where

$$\begin{aligned}
\frac{\partial L(\theta; y_i)}{\partial \beta} &= L^{-1}(\theta; y_i) \int L(\theta, b_i; y_i) \sum_{i=1}^{n_i} \left(y_{it} - p_{it}^c(b_{it})\right) \frac{\partial \Delta_{it}}{\partial \beta} f(b_i) db_i, \\
\frac{\partial L(\theta; y_i)}{\partial \gamma} &= L^{-1}(\theta; y_i) \int L(\theta, b_i; y_i) \left\{ \sum_{t=1}^{n_i} \left(y_{it} - p_{it}^c(b_{it})\right) \frac{\partial \Delta_{it}}{\partial \gamma} - \sum_{t=1}^{n_i} \frac{e_{it}}{\sigma_{i,t}^2} \frac{\partial e_{it}}{\partial \gamma} \right\} f(b_i) db_i, \\
\frac{\partial L(\theta; y_i)}{\partial \lambda} &= L^{-1}(\theta; y_i) \int L(\theta, b_i; y_i) \left\{ \sum_{t=1}^{n_i} \left(y_{it} - p_{it}^c(b_{it})\right) \frac{\partial \Delta_{it}}{\partial \lambda} + \sum_{t=1}^{n_i} \left(\frac{e_{it}^2}{\sigma_{i,t}^2} - 1 \right) h_{it} \right\} f(b_i) db_i,
\end{aligned}$$

with

$$\frac{\partial e_{i1}}{\partial \gamma} = 0, \qquad \frac{\partial e_{it}}{\partial \gamma} = -\sum_{j=1}^{t-1} b_{ij} w_{i,tj}.$$

To compute the score vector and information matrix, we also need derivatives of Δ_{it} with respect to β, γ, and λ. They can be obtained from the relationship (2.4),

$$\begin{aligned}
\frac{\partial P_{it}^M}{\partial \beta} &= \int \frac{\partial P_{it}^c(b_{it})}{\partial \triangle_{it}} \frac{\partial \triangle_{it}}{\partial \beta} f(b_i) db_i, \\
\Rightarrow \quad \frac{\partial \triangle_{it}}{\partial \beta} &= \frac{P_{it}^M \left(1 - P_{it}^M\right) x_{it}}{\int P_{it}^c(b_{it}) \left(1 - P_{it}^c(b_{it})\right) f(b_i) db_i}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\frac{\partial \triangle_{it}}{\partial \gamma} &= \frac{\int P_{it}^c(b_{it}) \sum_{k=1}^{n_i} \frac{e_{ik}}{\sigma_{ik}^2} \frac{\partial e_{ik}}{\partial \gamma} f(b_i) db_i}{\int P_{it}^c(b_{it}) \left(1 - P_{it}^c(b_{it})\right) f(b_i) db_i}, \\
\frac{\partial \triangle_{it}}{\partial \lambda} &= -\frac{\int P_{it}^c(b_{it}) \sum_{k=1}^{n_i} \left(\frac{e_{ik}^2}{\sigma_{ik}^2} - 1 \right) h_{ik} f(b_i) db_i}{\int P_{it}^c(b_{it}) \left(1 - P_{it}^c(b_{it})\right) f(b_i) db_i}.
\end{aligned}$$

Acknowledgement

We would like to thank Dr. Jungbok Lee of Korea University for providing the data and for their help in data collection and clarifying some issues with data.

References

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

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). Modeling repeated count data subject to informative dropout, *Journal of Multivariate Analysis*, **100**, 2352–2363.

Daniels, M. J. and Zhao, Y. D. (2003). Modelling the random effects covariance matrix in longitudinal data, *Statistics in Medicine*, **22**, 1631–1647.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses, *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 generalized linear mixed models, *Biometrika*, **88**, 973–985.

Kim, J., Kim, E., Yi, H., Joo, S., Shin, K., Kim, J., Kimm, K., and Shin, C. (2006). Short-term incidence rate of hypertension in Korea middle-aged adults. *Journal of Hypertension*, **24**, 2177–2182.

Lee, K. and Daniels, M. J. (2007). A Class of Markov models for longitudinal ordinal data, *Biometrics*, **63**, 1060–1067.

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**, 1284–1300.

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). Modeling 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 structures in longitudinal studies, *Biometrika*, **90**, 239–244.

Pan, J. and MacKenzie, G. (2006). Regression models for covariance structures in longitudinal studies, *Statistical Modelling*, **6**, 43–57.

Pinheiro, J. D. and Bates, D. M. (1996). Unconstrained parameterizations for variance-covariance matrices, *Statistics and Computing*, **6**, 289–296.

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.

Pourahmadi, M. and Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal fata, *Biometrics*, **58**, 225–231.

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.

Received December 22, 2013; Revised January 21, 2014; Accepted February 12, 2014