

Negative binomial loglinear mixed models with general random effects covariance matrix

Youkyung Sung^a, Keunbaik Lee^{1,a}

^aDepartment of Statistics, Sungkyunkwan University, Korea

Abstract

Modeling of the random effects covariance matrix in generalized linear mixed models (GLMMs) is an issue in analysis of longitudinal categorical data because the covariance matrix can be high-dimensional and its estimate must satisfy positive-definiteness. To satisfy these constraints, we consider the autoregressive and moving average Cholesky decomposition (ARMACD) to model the covariance matrix. The ARMACD creates a more flexible decomposition of the covariance matrix that provides generalized autoregressive parameters, generalized moving average parameters, and innovation variances. In this paper, we analyze longitudinal count data with overdispersion using GLMMs. We propose negative binomial loglinear mixed models to analyze longitudinal count data and we also present modeling of the random effects covariance matrix using the ARMACD. Epilepsy data are analyzed using our proposed model.

Keywords: overdispersion, ARMA Cholesky decomposition, positive-definite, longitudinal count data, high dimensionality

1. Introduction

Subjects are repeatedly observed over time in longitudinal studies; therefore, longitudinal data have between-subject variation as well as serial correlation of repeated outcomes (within-subject variation) (Diggle *et al.*, 2002). These variations should be considered for correct estimation of covariate effects. In addition, the variations can be modeled through a covariance matrix.

Modeling of the random effects covariance matrix in generalized linear mixed models (GLMMs) is a big issue in analysis of longitudinal categorical data (Breslow and Clayton, 1993) because the covariance matrix can be high-dimensional and its estimate must satisfy positive-definiteness (Lee *et al.*, 2012, Lee *et al.*, 2017). To avoid these constraints, the random effects covariance is assumed to be a simple structured matrix such as AR(1). However, it can be a very strong assumption, and this can lead to bias in estimates. To remove these obstacles, joint mean-covariance modeling approaches have been proposed such as modified Cholesky decomposition (MCD) (Lee, 2013; Lee and Sung, 2014; Pan and MacKenzie, 2003; Pourahmadi, 1999, 2000), moving average Cholesky decomposition (MACD) (Lee and Yoo, 2014; Zhang and Leng, 2012) and autoregressive-moving average Cholesky decomposition (ARMACD) (Choi and Lee, 2017; Han and Lee, 2016; Lee *et al.*, 2017; Nam and Lee, 2017).

In the MCD, precision matrix is decomposed to generalized autoregressive parameters (GARPs) and innovation variances (IVs) which are modeled using linear and loglinear models, respectively

¹ Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: keunbaik@skku.edu

(Pourahmadi, 1999, 2000). Therefore, the number of unknown elements in the precision matrix is reduced and the IVs are positive. In the MACD, covariance matrix is decomposed into generalized moving average parameters (GMAPs) and IVs. Similar to the MCD, the unknown elements of the covariance matrix is reduced, and the GMAPs and IVs are modeled using linear and loglinear models, respectively. The ARMACD combines the two Cholesky decompositions to model the covariance matrix (Lee *et al.*, 2017). Using the ARMACD, the covariance matrix has an autoregressive moving average (ARMA) structure. The ARMACD is a more flexible decomposition of the covariance matrix and factors the covariance matrix to GARPs, GMAPs, and IVs. In the above decompositions, all positive IVs guarantee the positive-definiteness of the estimated covariance matrix. In this paper, we exploit the ARMACD to model random effects covariance matrix in the GLMMs to analyze longitudinal count data. The ARMACD also allows flexible nonstationary and heteroscedastic random effects covariance matrix that is positive definite.

In this paper, we analyze longitudinal count data with overdispersion using GLMMs. To analyze longitudinal count data, Poisson loglinear mixed models (PLMMs) are typically used (Choi and Lee, 2017). In the models, identical mean and variance are assumed (Agresti, 2002). However, the variance is significantly larger than the mean in many cases. The greater variability than predicted by the Poisson models reflects overdispersion (Molenberghs *et al.*, 2007). To solve the overdispersion problem, random effects in the PLMMs are considered to analyze longitudinal count data (Thall and Vail, 1990). However, this approach cannot accommodate the hierarchies of longitudinal structure, which typically are presented in longitudinal data. For longitudinal overdispersed count data especially including autoregressive structure, Jowaheer and Sutradhar (2002) proposed generalized estimating equations (GEE) for regression parameters and the overdispersion parameter. Booth *et al.* (2003) extended the negative binomial loglinear model using random effects in the linear predictor for longitudinal overdispersed count data. Molenberghs *et al.* (2007) proposed a GLMM for longitudinal count data using gamma and normal types random effects. The model produces the standard negative-binomial (gamma random effects) and Poisson-normal (normal random effects) models as a special cases. In this paper, we also proposed negative binomial loglinear mixed models with general random effects covariance matrix using the ARMACD.

The outline of the paper is as follows. In Section 2, we propose modelling of the random effects covariance matrix using ARMACD in negative binomial loglinear mixed model and calculate maximum likelihood estimation. In Section 3, we apply our proposed models in epilepsy data and compare Poisson case. Finally, we summarize and proposed future work in Section 4.

2. Negative binomial loglinear mixed model

We now propose negative binomial loglinear mixed models for longitudinal overdispersed count data using the ARMACD.

2.1. Proposed model

Let $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ be the response vector of longitudinal count data where y_{ij} is the count response for subject i ($i = 1, \dots, N$) at time j ($j = 1, \dots, n_i$). We also assume that the responses for different subjects are conditionally independent given random effects. We now assume the following model,

$$\begin{aligned} y_{ij} &\sim NB(\mu_{ij}(b_{ij}), v^{-1}), \\ \log(\mu_{ij}(b_{ij})) &= x_{ij}^T \beta + b_{ij}, \end{aligned}$$

where $\mu_{ij}(b_{ij}) = E(y_{ij}|x_{ij}, b_{ij})$, β is a coefficients vector of x_{ij} , $b_i = (b_{i1}, \dots, b_{in_i})^T \sim N(0, \Sigma_i)$ is a vector of the random effect in the model, and ν is overdispersion parameter such that $\text{var}(y_{ij}|x_{ij}, b_{ij}) = \mu_{ij}(b_{ij}) + \nu(\mu_{ij}(b_{ij}))^2$. Note that the distribution of y_{ij} conditional b_{ij} has a negative binomial distribution which accommodate the overdispersion indexed by ν , and that the conditional variance of y_{ij} is greater than the conditional mean. Therefore, our proposed model accommodates longitudinal overdispersed count data.

Using the ARMACD, we assume the ARMA structure of the random effects covariance matrix which was used in analysis of longitudinal binary data (Lee *et al.*, 2017) and longitudinal Poisson data (Choi and Lee, 2017). The decomposition is given by

$$b_{ij} = \sum_{k=1}^{j-1} \phi_{ijk} b_{ik} + \sum_{k=1}^{j-1} l_{ijk} e_{ik} + e_{ij}, \quad (2.1)$$

where $e_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T \stackrel{\text{indep.}}{\sim} N(0, D_i)$ with $D_i = \text{diag}(\sigma_{i1}^2, \dots, \sigma_{in_i}^2)$, ϕ_{ijk} are GARPs, l_{ijk} are GMAPs, and σ_{ij}^2 are IVs.

Using matrix form, (2.1) is reexpressed as follows

$$T_i b_i = L_i e_i, \quad (2.2)$$

where T_i is a unique lower triangular matrix having ones on its diagonal and $-\phi_{ijk}$ at its $(j, k)^{\text{th}}$ element for $j < k$. Similarly, L_i is also a unique lower triangular matrix having ones on its diagonal and l_{ijk} at its $(j, k)^{\text{th}}$ element for $j < k$. Taking variance in (2.2), we have

$$T_i \Sigma_i T_i^T = L_i D_i L_i^T. \quad (2.3)$$

Note that Σ_i is decomposed to the GARPs, the GMAPs, and the IVs which are constrained and interpretable parameters ϕ_{ijk} , l_{ijk} , and $\log \sigma_{ij}^2$. In addition, the parameters ϕ_{ijk} , l_{ijk} , and $\log \sigma_{ij}^2$ are modeled using the following linear and loglinear models:

$$\phi_{ijk} = w_{ijk}^T \alpha, \quad l_{ijk} = z_{ijk}^T \gamma, \quad \log \sigma_{ij}^2 = h_{ij}^T \lambda, \quad (2.4)$$

where α and γ are vectors of unknown dependence parameters, λ is the vector of unknown variance parameters, and w_{ijk} , z_{ijk} , and h_{ij} are time and/or subject-specific design vectors. We note that w_{ijk} and z_{ijk} are covariate design vectors controlling the time order of the model and the correlation between responses. As a result, the random effects covariance matrix can be nonstationary and heteroscedastic depending on covariates.

Note that IVs are always positive using the loglinear model in (2.4). Therefore, diagonal elements of D_i in (2.1) are all positive and this result guarantees that Σ_i is positive definite by Theorem 1 in Lee *et al.* (2017). We also know that the T and L in (2.2) are identifiable using following theorem:

Proposition 1. *Matrices T and L in (2.2) are unique when Σ is positive definite and the order of ARMA structure is determined.*

Proof: From result in Pourahmadi (2011), matrix T is unique. Now we claim that matrix L is unique given matrix T . Since Σ is positive definite, $T_i \Sigma_i T_i$ is also positive definite. Then $T \Sigma T^T$ can be decomposed by the standard Cholesky decomposition:

$$T \Sigma T^T = C C^T, \quad (2.5)$$

where $C = (c_{ij})$ is a unique lower triangular matrix. Let $A = \text{diag}(c_{11}, \dots, c_{TT})$. Then (2.5) can be reexpressed by

$$\begin{aligned} T\Sigma T^T &= CA^{-1}AAA^{-1}C^T \\ &= LAAL^T \\ &= LDL^T, \end{aligned}$$

where $L = CA^{-1}$ and L is a unique lower triangular matrix, and $D = AA^T$. Thus, L are uniquely defined given T . \square

When the series is partly autoregressive and partly moving average, the ARMA models are parsimonious compared to the AR or the MA models (Judge *et al.*, 1980). Similarly, the linear and loglinear models for the GARPS/GMAPs and IVs in the ARMAD enable reasonable interpretation, easy computation, and stable estimation of the parameters (Lee *et al.*, 2017).

2.2. Maximum likelihood estimation

In this section, we derive maximum likelihood estimate of parameters for the proposed model using the likelihood function. Let $\theta = (\beta, \nu, \alpha, \gamma, \lambda)^T$ where θ consists of coefficients of covariates, dispersion parameter, and parameters of random effects covariance matrix. Then the marginal likelihood function is the integral of a product of negative binomial densities over random effects,

$$L(\theta; y) = \prod_{i=1}^N \int L(\theta; y_i, b_i) f(b_i) db_i, \quad (2.6)$$

where $f(b_i)$ is the multivariate normal distribution with mean 0 and variance covariance matrix Σ_i and

$$L(\theta; y_i, b_i) = \prod_{j=1}^{n_i} \frac{\Gamma(\nu^{-1} + y_{ij})}{\Gamma(\nu^{-1}) y_{ij}!} \left(\frac{1}{1 + \nu \mu_{ij}(b_{ij})} \right)^{\nu^{-1}} \left(\frac{\nu \mu_{ij}(b_{ij})}{1 + \nu \mu_{ij}(b_{ij})} \right)^{y_{ij}}.$$

Log-likelihood function is calculated as

$$\begin{aligned} \log L(\theta; y) &= \sum_{i=1}^N \log \int \exp \left[\sum_{j=1}^{n_i} \left\{ \sum_{l=0}^{y_{ij}-1} \log(\nu^{-1} + l) - \sum_{l=0}^{y_{ij}-1} \log(l + 1) \right. \right. \\ &\quad \left. \left. + y_{ij} \log \nu + y_{ij} \log \mu_{ij}(b_{ij}) - (\nu^{-1} + y_{ij}) \log(1 + \nu \mu_{ij}(b_{ij})) \right\} \right] f(b_i) db_i. \end{aligned}$$

We factor Σ_i using ARMA Cholesky decomposition, then we can represent $f(b_i)$ as

$$f(b_i) = (2\pi)^{-\frac{n_i}{2}} \left(\prod_{j=1}^{n_i} \sigma_{ij}^2 \right)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} b_i^T T_i^T L_i^{-T} D_i^{-1} L_i^{-1} T_i b_i \right).$$

Maximum Likelihood estimator is derived by solving the following equations:

$$\frac{\partial \log L(\theta; y)}{\partial \beta} = \sum_{i=1}^N \frac{1}{L(\theta; y_i)} \int L(\theta; y_i, b_i) \left[\sum_{j=1}^{n_i} \left\{ (y_{ij} - \mu_{ij}(b_{ij})) \frac{x_{ij}}{1 + \nu \mu_{ij}(b_{ij})} \right\} \right] f(b_i) db_i, \quad (2.7)$$

$$\frac{\partial \log L(\theta; y)}{\partial \alpha_l} = \sum_{i=1}^N \frac{1}{L(\theta; y_i)} \int L(\theta; y_i, b_i) f(b_i) \left\{ -\frac{1}{2} b_i^T \left(\frac{\partial T_i^T}{\partial \alpha_l} L_i^{-T} D_i^{-1} L_i^{-1} T_i + T_i^T L_i^{-T} D_i^{-1} L_i^{-1} \frac{\partial T_i}{\partial \alpha_l} \right) b_i \right\} db_i, \quad (2.8)$$

$$\frac{\partial \log L(\theta; y)}{\partial \gamma_l} = \sum_{i=1}^N \frac{1}{L(\theta; y_i)} \int L(\theta; y_i, b_i) f(b_i) \left\{ -\frac{1}{2} b_i^T \left(T_i^T \frac{\partial L_i^{-T}}{\partial \gamma_l} D_i^{-1} L_i^{-1} T_i + T_i^T L_i^{-T} D_i^{-1} \frac{\partial L_i^{-1}}{\partial \gamma_l} T_i \right) b_i \right\} db_i, \quad (2.9)$$

$$\frac{\partial \log L(\theta; y)}{\partial \lambda_l} = \sum_{i=1}^N \frac{1}{L(\theta; y_i)} \int L(\theta; y_i, b_i) \left\{ -\frac{1}{2} \sum_{j=1}^{n_i} h_{ijl} - \frac{1}{2} b_i^T T_i^T L_i^{-T} \frac{\partial D_i^{-1}}{\partial \lambda_l} L_i^{-1} T_i b_i \right\} \times f(b_i) db_i, \quad (2.10)$$

$$\frac{\partial \log L(\theta; y)}{\partial \nu} = \sum_{i=1}^N \frac{1}{L(\theta; y_i)} \int L(\theta; y_i, b_i) \left[\sum_{j=1}^{n_i} \left\{ \sum_{l=0}^{b_{ij}-1} -\frac{1}{\nu(1+\nu l)} + \frac{1}{\nu^2} \log(1+\nu \mu_{ij}(b_{ij})) \frac{y_{ij} - \mu_{ij}(b_{ij})}{\nu(1+\nu \mu_{ij}(b_{ij}))} \right\} \right] f(b_i) db_i, \quad (2.11)$$

where $L(\theta; y_i) = \int L(\theta; y_i, b_i) f(b_i) db_i$,

$$\frac{\partial T_i}{\partial \alpha_a} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -w_{i21a} & 0 & 0 & \cdots & 0 \\ -w_{i31a} & -w_{i32a} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -w_{in_1a} & -w_{in_2a} & -w_{in_3a} & \cdots & 0 \end{pmatrix}, \quad \frac{\partial L_i}{\partial \gamma_a} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ z_{i21,a} & 0 & 0 & \cdots & 0 \\ z_{i31,a} & z_{i32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{in_1,a} & z_{in_2,a} & z_{in_3,a} & \cdots & 0 \end{pmatrix},$$

$$\frac{\partial D_i^{-1}}{\partial \lambda_l} = \text{diag} \left\{ -\frac{1}{\sigma_{il}^2} h_{i1l}, \dots, -\frac{1}{\sigma_{in_l}^2} h_{in_l l} \right\}.$$

We can solve numerically the estimating equations (2.7)–(2.11) using quasi-Newton algorithm. The form is represented by following equations.

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

where $H(\theta^{(c)}; y)$ is a consistent and empirical estimator of the information matrix after convergence. It is given by

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

When this algorithm converged, the inverse of $H(\theta; y)$ can be the large-sample variance-covariance matrix of the parameter estimates.

Note that the solution of the equations (2.7)–(2.11) is computationally intensive because the high-dimensional random effects are integrated out. Quasi-Monte Carlo (QMC) approximation is used to compute integrations of the high-dimensional random effects in (2.6) (Lee *et al.*, 2012; Niederreiter, 1992; Pan and Thompson, 2007). The QMC works like the regular Monte Carlo method but instead of using a uniformly and randomly distributed set of points, uniformly distributed deterministic sequences, called low discrepancy sequences, are utilized. The QMC method seems to perform

better in high-dimensional random effects than either Gauss Hermite or Monte Carlo methods. The approximation of (2.6) is given by

$$L(\theta : y) \approx \prod_{i=1}^N \frac{1}{M} \sum_{l=1}^M \prod_{j=1}^{n_i} \frac{\Gamma(v^{-1} + y_{ij})}{\Gamma(v^{-1}) y_{ij}!} \left(\frac{1}{1 + v\mu_{ij}(b_{ij}^{(l)})} \right)^{v^{-1}} \left(\frac{v\mu_{ij}(b_{ij}^{(l)})}{1 + v\mu_{ij}(b_{ij}^{(l)})} \right)^{y_{ij}}, \quad (2.12)$$

where the set $(b_i^{(1)}, \dots, b_i^{(M)})$ is a subsequence of a low-discrepancy sequence with sample size of K . In Section 3, we used an R function, **rnorm.sobol()**, which is one of R function to make a low-discrepancy, in the library fOptions (Wuertz, 2005) to get the set. The QMC in (2.7)–(2.11) are also used to the approximation in (2.12).

3. Real data analysis

3.1. Data description

We next analyzed a data set from epilepsy study using our proposed model. The data set was first reported in a paper by Faught *et al.* (1996) and was analyzed in Molenberghs and Verbeke (2005) and Choi and Lee (2017). The study was a prospective open-label randomized non-comparative parallel study to compare placebo and a new anti-epileptic drug (AED), in combination with one or two other AED's. To stabilize the effect of AED, the number of seizures was collected for 12-week baseline period. A total 89 subjects with epilepsy were randomly assigned to one of the two groups (45 subjects for placebo and 44 subjects for AED treatment groups). The number of seizures was again counted over 16 weeks. The objective of this study was if the new AED has an effect on reducing the number of epileptic seizures.

Figure 1 shows average seizures for the two groups over week. There were no differences of average seizures between two groups until eighteen weeks. In week 19, there was extreme value in the placebo group, which is because very few observations were available.

We let response variable Y be the number of seizure and we included type of treatment ($\text{Trt} = 0$ for placebo group, $\text{Trt} = 1$ for AED group) to examine the effect of AED. The week number was standardized $((\text{week} - 14)/27)$ and interaction effect between treatment and week number was also included.

3.2. Model fit

We analyzed epilepsy study data using 4 Poisson linear mixed models (PLMMs) (Choi and Lee, 2017) and 6 negative binomial loglinear mixed models (NBLMMs) specified in Table 1.

We use the following notation: P-ARMA($?, ?$) and NB-ARMA($?, ?$) where P and NB respectively indicate Poisson linear mixed and negative binomial loglinear mixed models, and the ' $??$ ' correspond to the polynomial in the GARP and GMAP respectively or the AR and MA order, respectively. So an NB-ARMA(1, 1) would correspond to a heteroscedastic ARMA(1, 1) model with IV depending on type of treatment and NB-ARMA(L, L) would correspond to a heteroscedastic ARMA model with a linear in time difference for the GARP and GMAP. To find MLEs of each parameters, we used a QMC method that is explained in Subsection 2.2. The quasi-Newton algorithm was iterated until sum of absolute differences is less than 10^{-4} .

We compared the ten models using maximized loglikelihoods, Akaike information criteria (AIC) and Bayesian information criteria (BIC), presented in Table 2, respectively. Overall, negative binomial loglinear mixed models had much lower AICs and BICs than Poisson models. Therefore, we focused on the negative binomial loglinear mixed models.

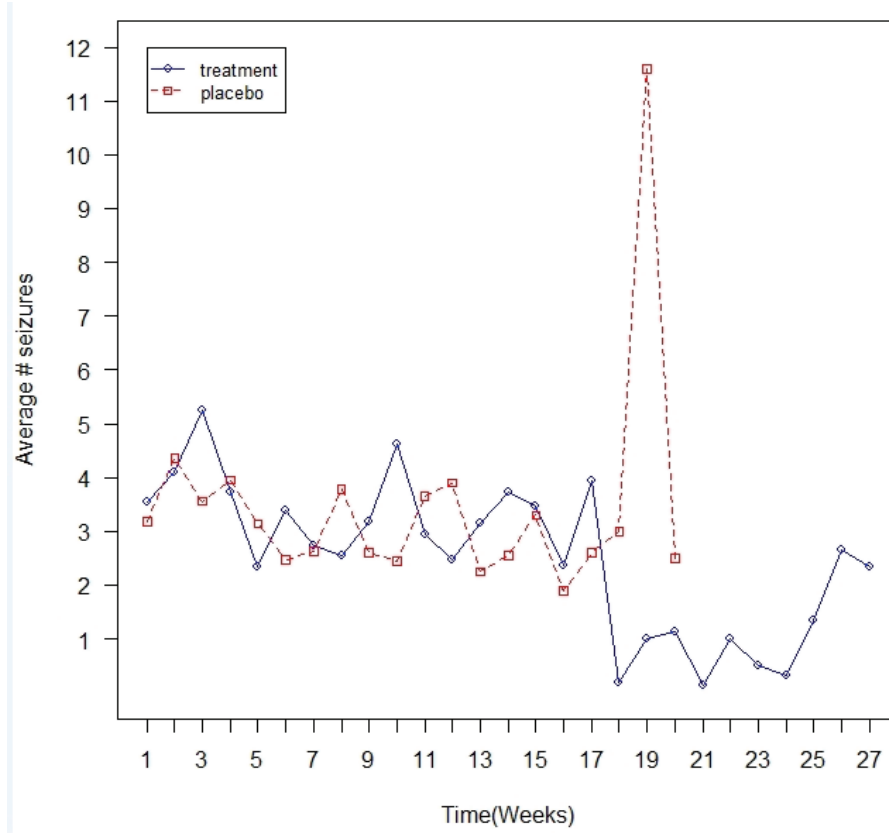


Figure 1: Average of number of seizure changes over time.

Table 1: Models for ϕ_{itj} , l_{itj} , and $\log \sigma_{it}^2$ for 4 PLMMs and 6 NBLMMs

Model		GARP	GMAP	IV
Poisson model	P-ARMA(1, 0)	$\phi_{itj} = \alpha_0 I_{(t-j)=1}$		$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	P-ARMA(2, 0)	$\phi_{itj} = \alpha_0 I_{(t-j)=1} + \alpha_1 I_{(t-j)=2}$		$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	P-ARMA(1, 1)	$\phi_{itj} = \alpha_0 I_{(t-j)=1}$	$l_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	P-ARMA(2, 1)	$\phi_{itj} = \alpha_0 I_{(t-j)=1} + \alpha_1 I_{(t-j)=2}$	$l_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
Negative binomial model	NB-ARMA(1, 0)	$\phi_{itj} = \alpha_0 I_{(t-j)=1}$		$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	NB-ARMA(2, 0)	$\phi_{itj} = \alpha_0 I_{(t-j)=1} + \alpha_1 I_{(t-j)=2}$		$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	NB-ARMA(1, 1)	$\phi_{itj} = \alpha_0 I_{(t-j)=1}$	$l_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	NB-ARMA(2, 1)	$\phi_{itj} = \alpha_0 I_{(t-j)=1} + \alpha_1 I_{(t-j)=2}$	$l_{itj} = \gamma_0 I_{(t-j)=1}$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	NB-ARMA(2, 2)	$\phi_{itj} = \alpha_0 I_{(t-j)=1} + \alpha_1 I_{(t-j)=2}$	$l_{itj} = \gamma_0 I_{(t-j)=1} + \gamma_1 I_{(t-j)=2}$	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$
	NB-ARMA(L, L)	$\phi_{itj} = \alpha_0 + \alpha_1 w_{ij} - w_{ik} $	$l_{itj} = \gamma_0 + \gamma_1 w_{ij} - w_{ik} $	$\log \sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Trt}_i$

PLMMs = Poisson loglinear mixed models; NBLMMs = negative binomial loglinear mixed models; GARPs = generalized autoregressive parameter; GMAP = generalized moving average parameter; IV = innovation variance; ARMA = autoregressive moving average.

We compared nested models (NB-ARMA(1,0) vs NB-ARMA(2,0), NB-ARMA(1,0) vs NB-ARMA(1,1), NB-ARMA(1,1) vs NB-ARMA(2,1), NB-ARMA(2,0) vs NB-ARMA(2,1), and NB-ARMA(2,1) vs NB-ARMA(2,2)) using likelihood ratio test (Table 3). It indicates that NB-ARMA(2,1) was the best model among above models. We also compared the above models and NB-ARMA(L, L)

Table 2: Log likelihood, AIC and BIC for the models

		P-ARMA(1, 0)	P-ARMA(2, 0)	P-ARMA(1, 1)	P-ARMA(2, 1)
Poisson	Max. loglik.	-2945.865	-2940.188	-2911.220	-2892.202
	AIC	5905.730	5896.376	5838.440	5802.404
	BIC	5923.150	5916.285	5858.349	5824.802
		NB-ARMA(1, 0)	NB-ARMA(2, 0)	NB-ARMA(1, 1)	NB-ARMA(2, 1)
Negative binomial	Max. loglik.	-2768.272	-2765.395	-2763.373	-2759.630
	AIC	5552.544	5548.790	5544.746	5539.260
	BIC	5572.453	5571.188	5567.144	5564.146
		NB-ARMA(2, 2)	NB-ARMA(L, L)		
Negative binomial	Max. loglik.	-2758.204	-2758.197		
	AIC	5538.408	5538.394		
	BIC	5565.783	5565.769		

AIC = Akaike information criteria; BIC = Bayesian information criteria; ARMA = autoregressive moving average.

Table 3: Likelihood ratio tests for negative binomial models

Models	χ^2	df	p-value
NB-ARMA(1, 0) vs NB-ARMA(2, 0)	5.754	1	0.017
NB-ARMA(1, 0) vs NB-ARMA(1, 1)	9.798	1	0.002
NB-ARMA(1, 1) vs NB-ARMA(2, 1)	7.486	1	0.006
NB-ARMA(2, 0) vs NB-ARMA(2, 1)	11.530	1	0.001
NB-ARMA(2, 1) vs NB-ARMA(2, 2)	2.852	1	0.091

df = degrees of freedom; ARMA = autoregressive moving average.

Table 4: Maximum likelihood estimates for NB-ARMA(2, 1), NB-ARMA(2, 2), and NB-ARMA(L, L) with standard errors in the parentheses

	NB-ARMA(2, 1)	NB-ARMA(2, 2)	NB-ARMA(L, L)
Mean parameter			
Intercept (β_0)	0.222 (0.326)	0.197 (0.421)	0.196 (0.357)
Treatment (β_1)	0.368 (0.351)	0.379 (0.376)	0.379 (0.401)
Week (β_2)	-1.582 (0.839)	-1.656 (1.148)	-1.657 (0.938)
Week×Treatment (β_3)	1.166 (0.854)	1.221 (0.962)	1.222 (1.017)
ν	0.278* (0.025)	0.275* (0.034)	0.275* (0.024)
GARP/GMAP			
α_0	1.512* (0.413)	1.172* (0.434)	-0.839* (0.425)
α_1	-0.506 (0.410)	-0.163 (0.428)	0.255 (0.538)
γ_0	-0.731 (0.292)	-0.250 (0.535)	1.270* (0.210)
γ_1		-0.269 (0.235)	13.012 (9.686)
IV			
λ_0	-1.597* (0.365)	-1.669* (0.553)	-1.669* (0.401)
λ_1	0.499 (0.267)	0.510 (0.367)	0.510 (0.727)

* indicates significance with 95% confidence level.

ARMA = autoregressive moving average; GARPs = generalized autoregressive parameter; GMAP = generalized moving average parameter; IV = innovation variance.

using AIC and BIC which are presented in Table 2. It indicates NB-ARMA(2, 1), NB-ARMA(2, 2), and NB-ARMA(L, L) were competitive. Therefore, we focused on the three models.

Table 4 provides maximum likelihood estimates of the parameters for NB-ARMA(2, 1), NB-ARMA(2, 2), and NB-ARMA(L, L). The parameters of coefficients of covariates in the three models were not significant. Thus, treatment did not appear to differentially impact the mean between placebo and treatment groups. In NB-ARMA(2, 1), overdispersion parameter was significant. It indicates that data were overdispersed compared to mean ($\widehat{\text{var}} = \hat{\mu}(1 + 0.278\hat{\mu})$ in NB-ARMA(2, 1)). The

AR(1) parameters in the GARPs and the intercept parameters in the IVs were significant ($\hat{\alpha}_0 = 1.512$, $SE = 0.413$; $\hat{\lambda}_0 = -1.597$, $SE = 0.365$). However, the AR(2) and MA(1) parameters were not significant. Also, the coefficients of treatment in the IVs was not significant. In NB-ARMA(2, 2), the parameter estimates were similar to those in NB-ARMA(2, 1). In NB-ARMA(L, L), the intercept parameters in the GARPs and GMAPs were significant. However, the coefficients of treatment were not significant. Similarly, the intercept parameter in the IVs was significant and the coefficient of treatment in the IVs was not significant.

4. Conclusion

We have proposed negative binomial loglinear mixed models to analyze longitudinal count data. The models account for the within-subject variation and between-subject variance using an ARMA random effects covariance matrix. The ARMA Cholesky decomposition (ARMACD) factors the random effects covariance matrix to GARPs, GMAPs and IVs. The GARPs, GMAPs, and IVs are modeled using linear and loglinear models which solve the positive definiteness constraint of the covariance matrix. In addition, the ARMACD creates a more flexible decomposition of the covariance matrix than existing decompositions. The estimation of parameters in the proposed models were calculated using a quasi-Newton algorithm. Calculation of the integrations in likelihood and derivatives was conducted by quasi Monte Carlo integration.

Epileptic seizure data were analyzed by Poisson/negative binomial loglinear mixed models with several structures of the random effects covariance matrix. The random effects covariance matrices with an ARMA(2, 1), an ARMA(2, 2), and an ARMA linear in time difference were competitive. The result shows that the parameters of coefficients of covariates in the models were not significant.

Acknowledgements

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-2016R1D1A1B03930343).

References

- Agresti A (2002). *Categorical Data Analysis* (2nd ed), Wiley and Sons, New York.
- Booth JG, Casella G, Friedl H, and Hobert JP (2003). Negative binomial loglinear mixed models, *Statistical Modelling*, **3**, 179–191.
- Breslow NE and Clayton DG (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 125–134.
- Choi J and Lee K (2017). Poisson linear mixed models with ARMA random effects covariance matrix, *Journal of the Korean Data & Information Science Society*, **28**, 659–668.
- Diggle PJ, Heagerty P, Liang KY, and Zeger S (2002). *Analysis of Longitudinal Data* (2nd ed), Oxford University Press, Oxford.
- Faught E, Wilder BJ, Ramsay RE, Reife RA, Kramer LD, Pledger GW, and Karim RM (1996). Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages, *Neurology*, **46**, 1684–1690.
- Han EJ and Lee K (2016). Dynamic linear mixed models with ARMA covariance matrix, *Communications for Statistical Applications and Methods*, **23**, 575–585.
- Jowaheer V and Sutradhar BC (2002). Analysing longitudinal count data with overdispersion, *Biomet-*

- rika*, **89**, 389–399.
- Judge GG, Griffiths WE, Hill RC, and Lee TC (1980). *The Theory and Practice of Econometrics*, Wiley, New York.
- Lee K (2013). Bayesian modeling of random effects covariance matrix for generalized linear mixed models, *Communications for Statistical Applications and Methods*, **20**, 235–240.
- Lee K, Baek C, and Daniels MJ (2017). ARMA Cholesky factor models for longitudinal regression models, *Computational Statistics & Data Analysis*, **115**, 267–280.
- Lee K, Lee J, Hagan J, and Yoo JK (2012). Modeling the random effects covariance matrix for the generalized linear mixed models, *Computational Statistics & Data Analysis*, **56**, 1545–1551.
- Lee K and Sung S (2014). Autoregressive Cholesky factor model for marginalized random effects model, *Communications for Statistical Applications and Methods*, **21**, 169–181.
- Lee K and Yoo JK (2014). Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models, *Computational Statistics and Data Analysis*, **80**, 111–116.
- Molenberghs G and Verbeke G (2005). *Models for Discrete Longitudinal Data*, Springer, New York.
- Molenberghs G, Verbeke G, and Demetrio CGB (2007). An extended random-effects approach to modeling repeated, overdispersed count data, *Lifetime Data Analysis*, **13** 513–531.
- Nam S and Lee K (2017). Comparison of the covariance matrix for general linear model, *The Korean Journal of Applied Statistics*, **30**, 103–117.
- Niederreiter H (1992). *Random Number Generation and Quasi-Monte Carlo Methods*, Siam, Philadelphia, Pennsylvania.
- Pan J and MacKenzie G (2003). On modelling mean-covariance structures in longitudinal studies, *Biometrika*, **90**, 239–244.
- Pan J and Thompson R (2007). Quasi-Monte Carlo estimation in generalized linear mixed models, *Computational Statistics & Data Analysis*, **51**, 5765–5775.
- 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. (2011). Covariance estimation: The GLM and regularization perspectives, *Statistical Science*, **26**, 369–387.
- Thall PF and Vail SC (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics*, **46**, 657–671.
- Wuertz D (2005). fOptions: Financial Software Collection-fOptions. R package version 220.10063. (<http://www.rmetrics.org>).
- Zhang W and Leng C (2012). A moving average Cholesky factor model in covariance modelling for longitudinal data, *Biometrika*, **99**, 141–150.