Communications for Statistical Applications and Methods
2020, Vol. 27, No. 2, 201–210
https://doi.org/10.29220/CSAM.2020.27.2.201
Print ISSN 2287-7843 / Online ISSN 2383-4757

Bayesian baseline-category logit random effects models for longitudinal nominal data

Jiyeong Kim[a] , Keunbaik Lee[1,b]

[a]Laboratory of Low Dose Risk Assessment, National Radiation Emergency Medical Center, Korea Institute of Radiological and Medical Sciences, Korea;
[b]Department of Statistics, Sungkyunkwan University, Korea

Abstract

Baseline-category logit random effects models have been used to analyze longitudinal nominal data. The models account for subject-specific variations using random effects. However, the random effects covariance matrix in the models needs to explain subject-specific variations as well as serial correlations for nominal outcomes. In order to satisfy them, the covariance matrix must be heterogeneous and high-dimensional. However, it is difficult to estimate the random effects covariance matrix due to its high dimensionality and positive-definiteness. In this paper, we exploit the modified Cholesky decomposition to estimate the high-dimensional heterogeneous random effects covariance matrix. Bayesian methodology is proposed to estimate parameters of interest. The proposed methods are illustrated with real data from the McKinney Homeless Research Project.

Keywords: covariance matrix, heterogeneous, high-dimensional, modified Cholesky decomposition, positive-definiteness

1. Introduction

Longitudinal data are collected over time from the same subjects. Therefore, the outcomes from the same subjects are correlated. Many models have been proposed to analyze data such as linear mixed models and generalized linear mixed models (GLMMs). Especially, the GLMMs are commonly used to analyze longitudinal categorical data (Breslow and Clayton, 1993), and the GLMMs specify the effects of covariates on response conditional for random effects.

Baseline-category logit random effects models are typically used to analyze longitudinal nominal data (Theil, 1969, 1970), and the models account for subject-specific variations using the random effects covariance matrix. However, the random effects covariance matrix in the models cannot explain the serial correlations of nominal outcomes. The random effects covariance matrix must be heterogeneous and high-dimensional to account for both the correlations and subject-specific variations. However, it is difficult to estimate the random effects covariance matrix due to high dimensionality and positive-definiteness (Lee *et al.*, 2012). In this paper, we propose models to solve problems using modified Cholesky decomposition (MCD) (Pourahmadi, 1999).

The MCD approach uses the new unconstrained parametrization of an inverse of a covariance matrix. As a result, the parameters of the MCD are generalized autoregressive parameters (GARPs)

This paper is based on part of Jiyeong Kim's PhD thesis.

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

Published 31 March 2020 / journal homepage: http://csam.or.kr

and innovation variances (IVs). The GARPs are dependence parameters describing the serially dependence of the previous outcomes, and the IVs are prediction variances. The positive-definiteness restriction of the covariance matrix is that the IVs need to be positive (Pourahmadi, 1999, 2000). Pan and Mackenzie (2006) used the MCD to address joint mean-covariance estimation for linear models. Lee *et al.* (2012) and Lee (2013) also used the MCD to estimate the random effects covariance matrix in logistic random effects models for longitudinal binary data. In this paper, we also use the MCD to estimate the random effects covariance matrix in baseline-category logit random effects models for longitudinal nominal data.

There is much literature dealing with models for longitudinal nominal data. Multinomial logit models were developed by Theil (1969, 1970). Daniels and Gatsonis (1997) proposed a Bayesian two-level generalized logit model to accommodate clustered nominal data. Revelt and Train (1998) proposed discrete choice models with random coefficients that do not have the restrictive 'independence from irrelevant alternatives' property. Hartzel *et al.* (2001) studied logit random effects models of clustered ordinal and nominal data. For the analysis of clustered or longitudinal nominal data, mixed-effects multinomial logistic regression models were proposed by Hedeker (2003). Chen *et al.* (2009) developed a Markov model based on the likelihood approach to analyze longitudinal categorical data for the modeling of both marginal and conditional relationships. Lee and Mercante (2010) and Lee *et al.* (2011) proposed marginalized models using a Markovian dependence structure or random effects to analyze longitudinal nominal data, respectively.

This paper is organized as follows. In Section 2, we propose baseline-category logit random effects models for longitudinal nominal data using the MCD approach. In Section 3, we present Bayesian methodology for estimation of parameters. In Section 4, we illustrate real data and apply our proposed models to them. Finally, we summarize this paper in Section 5.

2. Bayesian baseline-category logit random effects models for longitudinal nominal data

In this section, we propose baseline-category logit random effects models with autoregressive random effects covariance matrix to analyze longitudinal nominal data.

2.1. Proposed models

Let Y_{it} be a nominal response with K-categories on subject i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, n_i$; $n_i \leq T$) and let x_{it} be the corresponding vector of covariates. We assume that each Y_{it} is conditionally independent given random effects b_{it}, that the responses for different subjects are independent, and that the regression model is given by

$$\log \frac{P(Y_{it} = k | b_{it}, x_{it})}{P(Y_{it} = K | b_{it}, x_{it})} = x_{it}^T \beta_k + b_{itk}, \tag{2.1}$$

where $k = 1, \ldots, K-1$, x_{it} is a $p \times 1$ vector of covariates, β_k is a $p \times 1$ vector of regression coefficient,

$$b_i = (b_{i1}, \ldots, b_{in_i})^T \overset{\text{indep}}{\sim} N(0, \Sigma_i), \tag{2.2}$$

with $b_{it}^T = (b_{it1}, \ldots, b_{it,K-1})$, and Σ_i is a $(K-1)n_i \times (K-1)n_i$ random effects covariance matrix.

Then conditional probabilities for Y_{it} given the random effects b_{it} are given by

$$P(Y_{it}=k|b_{it},x_{it})=\begin{cases}\dfrac{\exp\left(x_{it}^T\beta_k+b_{itk}\right)}{1+\sum_{l=1}^{K-1}\exp\left(x_{it}^T\beta_k+b_{itk}\right)}, & \text{for } k=1,\ldots,K-1,\\ \dfrac{1}{1+\sum_{l=1}^{K-1}\exp\left(x_{it}^T\beta_k+b_{itk}\right)}, & \text{for } k=K.\end{cases}$$

The random effects covariance matrix Σ_i has subject variations and serial correlations. In addition, it is high-dimensional and should be positive definite. However, the estimation of the covariance matrix is not easy due to the constraints. Therefore, we consider the MCD to solve the problem of constraints.

2.2. Modeling of the random effects covariance matrix

In this section, we describe the random effects covariance matrix Σ_i using the MCD. The random effects b_{it} in equation (2.2) are assumed to be decomposed as follows

$$b_{i1}=e_{i1}, \tag{2.3}$$

$$b_{it}=\sum_{j=1}^{t-1}\Psi_{itj}b_{ij}+e_{it},\quad \text{for } t=2,\ldots,n_i, \tag{2.4}$$

where

$$e_{it}=\begin{pmatrix}e_{it1}\\ \vdots\\ e_{it,K-1}\end{pmatrix},\quad \Psi_{itj}=\begin{pmatrix}\phi_{itj,11} & \phi_{itj,12} & \cdots & \phi_{itj,1,K-1}\\ \phi_{itj,21} & \phi_{itj,22} & \cdots & \phi_{itj,2,K-1}\\ \vdots & \vdots & \ddots & \vdots\\ \phi_{itj,K-1,1} & \phi_{itj,K-1,2} & \cdots & \phi_{itj,K-1,K-1}\end{pmatrix}.$$

Note that the elements of the matrix Ψ_{itj}, which are called GARPs, present serial correlations of repeated nominal outcomes. The serial correlations are the *correlation within categories over time* and the *cross-correlation between different categories at different times*. A similar form to indicate dependence of multivariate longitudinal outcomes was presented in Lee *et al.* (2020). We refer to the matrix Ψ_{itj} as the generalized autoregressive matrices (GARMs).

We also assume that

$$e_i=\left(e_{i1}^T,\ldots,e_{in_i}^T\right)^T \overset{\text{indep}}{\sim} N(0,A_i),$$

where $A_i=\text{diag}(A_{i1},\ldots,A_{in_i})$ with $A_{it}=\text{diag}(\sigma_{it1}^2,\ldots,\sigma_{it,K-1}^2)$. Note that diagonal matrix A_{it} presents the prediction variance matrix of b_{it}. We refer to the new parameters A_{it} as the innovation variance matrices (IVMs).

Then we reexpress equations (2.3) and (2.4) in matrix form as

$$T_ib_i=e_i, \tag{2.5}$$

where

$$T_i=\begin{pmatrix}\mathrm{I} & 0 & 0 & \ldots & 0\\ -[\Psi_{i21}] & \mathrm{I} & 0 & \ldots & 0\\ -[\Psi_{i31}] & -[\Psi_{i32}] & \mathrm{I} & \ldots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ -[\Psi_{in_i1}] & -[\Psi_{in_i2}] & -[\Psi_{in_i3}] & \ldots & \mathrm{I}\end{pmatrix}.$$

From equation (2.5), we have

$$T_i\Sigma_iT_i^T = \mathrm{Var}(e_i) \Longrightarrow T_i\Sigma_iT_i^T = A_i \\ \Longrightarrow \Sigma_i = T_i^{-1}A_iT_i^{-T}. \tag{2.6}$$

We note that the random effects covariance matrix is directly decomposed into the GARMs and IVMs. They can be modeled using time and/or subject-specific covariate vectors w_{itj} and h_{it} by setting

$$\phi_{itj,lm} = w_{itj}^T\alpha_{lm}, \quad \log\sigma^2_{itk} = h_{it}^T\lambda_k, \tag{2.7}$$

where α_{lm} is an $a \times 1$ vector and λ_k is a $b \times 1$ vector of unknown parameters, respectively.

Note that w_{itj}'s are covariate design vectors controlling the time order of the model and the correlation of responses. As a result, the random effects covariance matrix can be nonstationary and heteroscedastic depending on covariates. We also note that the IVMs are positive definite using the loglinear model in equation (2.7). This result guarantees that Σ_i is positive definite because the diagonal matrix of A_i in equation (2.6) are all positive. These results represent the advantages of the MCD.

3. Bayesian methodology

We now describe Bayesian approaches to estimate parameters in our proposed models. We derive the likelihood function for the model specified in Subsection 2.1. The parameters in model (2.7) are the regression coefficients which ranges on $(-\infty, \infty)$. In this case, normal priors are commonly used for parameters and guarantee the propriety of posterior distributions. The normal priors with large prior variances remains relatively objective (Daniels and Zhao, 2003). The priors distributions for the parameters in the model with the AR structure of random effects covariance matrix are given by

$$\beta_k \sim N\left(0, \sigma^2_\beta I\right), \tag{3.1}$$

$$\alpha_{lm} \sim N\left(0, \sigma^2_\alpha I\right), \tag{3.2}$$

$$\lambda_k \sim N\left(0, \sigma^2_\lambda I\right). \tag{3.3}$$

In general, σ^2_β, σ^2_α, and σ^2_λ are in large to be noninformative such as 100. By combining likelihood function from (2.1) and prior distributions from (3.1)–(3.3), we obtain the joint distribution given by

$$\begin{aligned} P(y, b, \beta, \alpha, \lambda) &\propto P(y|b, \beta)P(b|\alpha, \lambda)P(\beta)P(\alpha)P(\lambda) \\ &\propto \left[\prod_{i=1}^{N}\left\{\prod_{t=1}^{n_i}\prod_{k=1}^{K}\left(P^c_{itk}(b_{itk})\right)^{y_{itk}}\right\}\left\{\prod_{t=1}^{n_i}\prod_{k=1}^{K-1}(\sigma^2_{itk})^{-\frac{1}{2}}\right\}\exp\left(-\frac{1}{2}b_i^T\Sigma_i^{-1}b_i\right)\right] \\ &\quad \times \prod_{k=1}^{K-1}\exp\left(-\frac{1}{2\sigma^2_\beta}\beta_k^T\beta_k\right) \times \prod_{l=1}^{K-1}\prod_{m=1}^{K-1}\exp\left(-\frac{1}{2\sigma^2_\alpha}\alpha_{lm}^T\alpha_{lm}\right) \times \prod_{k=1}^{K-1}\exp\left(-\frac{1}{2\sigma^2_\lambda}\lambda_k^T\lambda_k\right). \end{aligned}$$

To generate parameters from the posterior distribution, MCMC methods are adapted to generate posterior samples for model estimation. The full condition posterior distribution are given below:

- For b_i $(i = 1, \ldots, N)$

$$P(b_i|y, \beta, \beta_0, \alpha, \gamma, \lambda) \propto \left\{\prod_{t=1}^{n_i}\prod_{k=1}^{K}\left(P^c_{itk}(b_{it})\right)^{y_{itk}}\right\} \times \exp\left(-\frac{1}{2}\sum_{i=1}^{N}b_i^TT_i^TA_i^{-1}T_i^Tb_i\right).$$

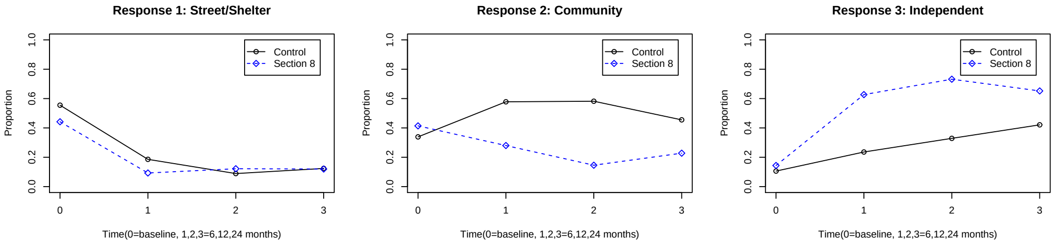

Figure 1: *Response of marginal proportions under two group (Section 8 and control) for response 1 (street/shelter), 2 (community) and 3 (independent housings), respectively.*

- For β_k

$$P(\beta_k|y, b, \beta_0, \alpha, \gamma, \lambda) \propto \left[\prod_{i=1}^{N}\left\{\prod_{t=1}^{n_i}\prod_{k=1}^{K}\left(P^c_{itk}(b_{it})\right)^{y_{itk}}\right\}\right]\exp\left(-\frac{1}{2\sigma^2_{\beta_k}}\beta_k^T\beta_k\right).$$

- For α_{lm}

$$P(\alpha_{lm}|y, b, \beta, \beta_0, \gamma, \lambda) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{N} b_i^T T_i^T A_i^{-1} T_i^T b_i\right)\exp\left(-\frac{1}{2\sigma^2_{\alpha}}\alpha^T\alpha\right).$$

- For λ_k

$$P(\lambda_k|y, b, \beta, \beta_0, \alpha, \gamma) \propto \left\{\prod_{i=1}^{N}\prod_{t=1}^{n_i}\left(\sigma^2_{it}\right)^{-\frac{1}{2}}\right\}\exp\left(-\frac{1}{2}\sum_{i=1}^{N} b_i^T T_i^T A_i^{-1} T_i^T b_i\right)\exp\left(-\frac{1}{2\sigma^2_{\lambda_k}}\lambda_k^T\lambda_k\right).$$

All full conditionals are not closed forms; therefore, we construct suitable proposals for a Metropolis-Hastings step. In practice, MCMC is implemented using JAGS (http://mcmc-jags.sourceforge.net/).

4. Analysis of McKinney homeless research project data

4.1. Data description

The McKinney homeless research project (MHRP), first described by Hurlburt *et al.* (1996), was used to demonstrate the use of our proposed models. The MHRP was designed to assess if the use of Section 8 housing certificates was effective in providing housing options to homeless individuals with severe mental illness. 361 clients were randomly assigned to one of two types of supportive case management (comprehensive and traditional) and to one of two levels of access for independent housing (Section 8 certificates: Group = 1 for yes; 0 for no). Nominal-level housing status outcomes were collected at baseline, and at 6, 12, and 24 months after post-randomization. When analyzing this data, time was set to $0, 1, 2, 3$. Similar to Hedeker and Gibbons (2006), we focus on examining the effect of access to Section 8 certificates on repeated housing outcomes across time. There were three different housing outcomes of either street/shelter housing, community housing or independent housing for each time point.

Figure 1 presents marginal proportions of responses for MHRP data. The marginal proportion plots indicate the marginal probability for Section 8 certificate at each time, indicating the difference between the groups.

Table 1: Nominal outcomes by living arrangements

Outcomes	Living arrangement
Street/shelter housing	Public or private shelter Church/chapel Indoor public place (bus station/theater) Abandoned building Car or other vehicle Outside without shelter
Community housing	Hotel Family member's home or room Friend's or acquaintance's home or room Boarding house/halfway house
Independent housing	Private house or own apartment

Table 2: Models fit with w_{itj} and h_{it} for MHRP data

Model	GARMs	log(ICMs)
GLMM	NA	$\log\sigma^2_{itk} = \lambda_{0k}$
AR(1)-CC:	$\phi_{itj,lm} = \alpha_{lm,0}I_{(\|t-j\|=1)}$	$\log\sigma^2_{itk} = \lambda_{0k}$
AR(1)-CA:	$\phi_{itj,lm} = \alpha_{lm,0}I_{(\|t-j\|=1)}$	$\log\sigma^2_{itk} = \lambda_{0k} + \lambda_{1k}\text{Group}_i$
AR(1)-AC:	$\phi_{itj,lm} = \alpha_{lm,0}I_{(\|t-j\|=1)} + \alpha_{lm,1}\text{Group}_i I_{(\|t-j\|=1)}$	$\log\sigma^2_{itk} = \lambda_{0k}$
AR(1)-AA:	$\phi_{itj,lm} = \alpha_{lm,0}I_{(\|t-j\|=1)} + \alpha_{lm,1}\text{Group}_i I_{(\|t-j\|=1)}$	$\log\sigma^2_{itk} = \lambda_{0k} + \lambda_{1k}\text{Group}_i$

For more detailed explanations on categorical outcomes, housing categories by living arrangements are summarized in Table 1 according to the study by Hurlburt *et al.* (1996). About 25 % of the subjects dropped out of the study during the follow-up period resulting in some missing housing outcome status data. For our analysis, we assume missing at random missingness. The housing outcome status of street/shelter was chosen as the reference category; in addition, we focused on evaluating the effects of access to Section 8 certificates on housing outcomes across time.

4.2. Model fit

We fit five models which are a typical GLMM and four baseline-category logit random effects models with several structure of Σ_i.

Table 2 presents the specification of the five models for Σ_i using the various structures. The typical GLMM is a baseline-category logit random effects model with a homogeneous random effects variance. The other models considered only the time window difference using indicator function $1(I_{(|t-j|=1)})$, which indicates the AR(1) structure. Note that '-C' means a model with a constant random effects covariance matrix and '-A' means a model with a random effect covariance matrix depending on group. Therefore, Model AR(1)-CC indicates a model with a homogeneous random effects covariance matrix with an AR(1) structure, and Model AR(1)-AA indicates a model with a group-dependent AR(1) random effects covariance matrix having IVs depending on a group with an AR(1) structure. Model AR(1)-CA is a model with an AR(1) random effects covariance matrix having an IVs depending on a group. Finally, Model AR(1)-AC is a model with a group-dependent AR(1) random effects covariance matrix that has constant IVs.

For the estimation of all parameters in the models, Gibbs sampler is implemented using JAGS in R. Posterior means were calculated with a sample size of 250,000, thin of 5 and burn-in period of 100,000. To use the Gelman and Rubin approach, we used multiple chains (chain of 2). We also checked the convergence of all parameters in the models using the trace plots of random numbers for the parameters. Using the plots, we observed that the lines of different chains were mixed and crossed;

Table 3: DIC of Bayesian baseline-category logit random effects model for MHRP data

Model	$\overline{\text{Dev}(\theta)}$	$D(\widetilde{\theta})$	p_D	DIC
GLMM	2069.00	1666.34	402.66	2471.66
AR(1)-CC	1612.81	1227.92	384.89	1997.70
AR(1)-CA	1629.63	1261.44	368.19	1997.82
AR(1)-AC	1590.25	1208.74	381.51	1971.76
AR(1)-AA	1526.97	1090.70	436.27	1963.24

Table 4: Posterior means of Bayesian baseline-category logit random effects model using MCD for MHRP data (95% Bayesian confidence interval)

	GLMM	AR(1)-CC	AR(1)-CA	AR(1)-AC	AR(1)-AA
C vs S					
Intercept	−0.09 (−0.50, 0.33)	−0.24 (−0.58, 0.09)	−0.24 (−0.60, 0.12)	−0.21 (−0.55, 0.12)	−0.27 (−1.34, 0.90)
6 Month vs Baseline	1.71* (0.74, 3.11)	1.62* (1.02, 2.30)	1.64* (1.04, 2.34)	1.61* (1.03, 2.26)	3.01* (1.40, 5.65)
12 Month vs Baseline	1.99* (0.68, 3.22)	2.60* (1.71, 3.62)	2.62* (1.71, 3.63)	2.73* (1.81, 3.79)	3.48* (1.37, 5.67)
24 Month vs Baseline	0.79 (−1.14, 1.96)	2.22* (0.80, 3.68)	2.26* (0.89, 3.68)	2.53* (1.02, 4.19)	2.48 (−0.30, 4.91)
Section 8 (YES, NO)	−0.88 (−3.33, 0.29)	0.07 (−0.47, 0.55)	0.05 (−0.57, 0.57)	0.00 (−0.51, 0.48)	−0.10 (−1.33, 0.88)
Section 8 by 6 Month	−1.34 (−4.23, 0.49)	−0.38 (−1.39, 0.59)	−0.37 (−1.43, 0.64)	−0.19 (−1.31, 0.93)	−1.34 (−4.09, 0.61)
Section 8 by 12 Month	−3.66 (−7.42, −1.20)	−2.42* (−3.71, −1.21)	−2.39* (−3.80, −1.09)	−2.19* (−3.88, −0.62)	−2.77* (−5.58, −0.31)
Section 8 by 24 Month	−1.45 (−4.20, 0.36)	−0.90 (−2.91, 0.81)	−0.72 (−2.86, 1.19)	−0.95 (−3.57, 1.44)	−0.54 (−3.66, 2.63)
λ_0	2.05 (−3.25, 4.62)	−1.64 (−5.69, 1.11)	−1.22 (−6.10, 1.46)	−1.46 (−5.54, 0.44)	3.40* (0.61, 6.14)
λ_1 (Group)			−0.07 (−1.81, 1.40)		−3.06* (−6.27, −0.27)
I vs S					
Intercept	−0.52* (-0.83,-0.21)	−1.47* (-2.02,-0.98)	−1.44* (−1.99, −0.96)	−1.62* (−2.35, −1.06)	−1.67* (−2.46, −1.08)
6 Month vs Baseline	0.98* (0.41,1.57)	1.52* (0.74,2.30)	1.58* (0.80, 2.34)	1.56* (0.74, 2.38)	1.94* (1.09, 2.89)
12 Month vs Baseline	1.97* (1.30,2.69)	2.52* (1.23,3.80)	2.64* (1.33, 3.87)	2.76* (1.38, 4.11)	3.36* (1.95, 5.02)
24 Month vs Baseline	1.66* (1.02,2.32)	2.07* (0.23,3.84)	2.24* (0.39, 3.95)	2.63* (0.61, 4.60)	2.83* (0.72, 5.15)
Section 8 (YES, NO)	−0.70* (-1.27,-0.16)	−0.02 (-0.70,0.64)	−0.13 (−0.92, 0.60)	0.08 (−0.63, 0.79)	−0.75 (−2.58, 0.62)
Section 8 by 6 Month	2.23* (1.29,3.23)	2.03* (0.89,3.26)	2.09* (0.88, 3.51)	2.25* (0.94, 3.81)	3.25* (1.15, 5.89)
Section 8 by 12 Month	0.86 (-0.11,1.84)	1.35 (-0.14,2.93)	1.39 (−0.18, 3.09)	1.73 (−0.05, 3.84)	2.70* (0.10, 5.70)
Section 8 by 24 Month	1.50* (0.52,2.52)	3.082* (0.91,5.41)	3.11* (0.77, 5.62)	2.95* (0.25, 5.82)	4.08* (0.88, 7.47)
λ_0	−2.92* (-6.67,-0.21)	−0.50 (-1.80,0.69)	−0.74 (−2.40, 0.65)	−0.26 (−1.91, 1.19)	−0.16 (−2.11, 1.45)
λ_1 (Group)			0.44 (−0.76, 1.46)		1.33 (−2.96, 3.73)
$\alpha_{11,0}$ (AR(1))		1.91* (0.70,4.40)	1.72* (0.61, 4.91)	1.78* (0.80, 4.62)	0.55* (0.32, 0.97)
$\alpha_{12,0}$ (AR(1))		0.05 (-0.18,0.25)	0.05 (−0.20, 0.26)	−0.52 (−2.36, 0.73)	0.02 (−1.28, 1.22)
$\alpha_{21,0}$ (AR(1))		−0.69 (-5.27,0.90)	−0.34 (−5.40, 0.98)	0.19 (−0.16, 0.52)	−0.15 (−1.32, 0.57)
$\alpha_{22,0}$ (AR(1))		2.11* (1.48,2.94)	2.11* (1.38, 3.09)	−0.18 (−0.63, 0.23)	0.17 (−0.62, 1.35)
$\alpha_{11,1}$ (AR(1) by Group)				−0.655 (−5.35, 0.88)	0.01 (−0.09, 0.21)
$\alpha_{12,1}$ (AR(1) by Group)				3.26 (−2.04, 8.48)	3.14* (0.18, 8.07)
$\alpha_{21,1}$ (AR(1) by Group)				1.94* (1.27, 2.93)	2.15* (1.33, 3.33)
$\alpha_{22,1}$ (AR(1) by Group)				−0.266 (−1.97, 0.89)	−0.95* (−2.34, 0.37)

* indicates the 95% credible interval does not include zero. C vs S is Community vs Street/Shelter, I vs S is Independent vs Street/Shelter

convergence was then satisfied.

Table 3 shows comparison of the four models using deviance information criterion (DIC) (Spiegelhalter *et al.*, 2010). Since AR(1)-AA was the most complex model among the five models, p_D was large. However, DIC was the smallest of all other models. This means that the Model AR(1)-AA was the best fit. Therefore, we now focus on Model AR(1)-AA for further analysis.

Table 4 is organized into three parts according to the nominal response categories to be compared (either community vs street/shelter or independent vs street/shelter) and GARPs. The top part of Table 4 presents the estimates of coefficients and associated 95% credible interval to compare the two nominal response categories of community housing and street/shelter housing. For these two categories, the estimated baseline-category logit is given by

$$\log \frac{\hat{P}(\text{Community})}{\hat{P}(\text{Street/Shelter})} = -0.27 + 3.01\text{Time}_1 + 3.48\text{Time}_2 + 2.48\text{Time}_3 - 0.10\text{Section 8}$$
$$- 1.34\text{Time}_1 * \text{Section 8} - 2.77\text{Time}_2 * \text{Section 8} - 0.54\text{Time}_3 * \text{Section 8} + \hat{b}_{itK},$$

where Time_1, Time_2, and Time_3 are indicators of month 6 (6 Month), month 12 (12 Month), and month 24 (14 Month), respectively. The posterior means of regression coefficients for 6 Month versus Baseline and 12 Month versus Baseline were not in 95% credible intervals. In Control (Group_i = 0), the odds ratio of the conditional posterior probability were $e^{3.01} = 20.29$, $e^{3.48} = 32.46$, and $e^{2.48} = 11.94$ in community housing as opposed to street/shelter housing at 6 Month, 12 Month and 24 Months. The posterior means of regression coefficients for the interaction between Section 8 certificates and 12 Month were not in the credible intervals. Combining the logit estimates for the main effects of Section 8 certificates (−0.10) and the association between 6 Month and Section 8 certificates (−1.34) yielded an estimated odds ratio of $e^{-1.44} = 0.24$, the association between 12 Month and Section 8 certificates (−2.77) yielded an estimated odds ratio of $e^{-2.87} = 0.06$, and the association between 24 Month and Section 8 certificates (−0.54) yielded an estimated odds ratio of $e^{-0.64} = 0.53$, suggesting that individuals with Section 8 certificates were less likely to be in community housing as opposed to street/shelter housing at 6 Month, 12 Month and 24 Months.

Second, to compare independent housing and street/shelter housing we consider the lower part of Table 4. The baseline-category logit is

$$\log\frac{\hat{P}(\text{Independent})}{\hat{P}(\text{Street/Shelter})} = -1.67 + 1.94\text{Time}_1 + 3.36\text{Time}_2 + 2.83\text{Time}_3 - 0.75\text{Section 8} \\ + 3.25\text{Time}_{10}*\text{Section 8} + 2.70\text{Time}_2*\text{Section 8} + 4.08\text{Time}_3*\text{Section 8} + \hat{b}_{itK}.$$

The posterior means of regression coefficients for 6 Month versus Baseline, 12 Month versus Baseline and 24 Month versus Baseline were not in credible intervals. The odds ratios of the control group were $e^{1.94} = 6.96$, $e^{3.36} = 28.79$, and $e^{2.83} = 16.95$ in community housing as opposed to street/shelter housing at 6 Month, 12 Month, and 24 Months. Also the posterior means of regression coefficients of the interaction between Section 8 certificates and 6 Month, 12 Month, and 24 Month were not in credible intervals. Combining the logit estimates for the main effects of Section 8 certificates (−0.75) and the association between 6 Month and Section 8 certificates (3.25) yielded an estimated odds ratio of $e^{2.5} = 12.18$, the association between 12 Month and Section 8 certificates (2.70) yielded an estimated odds ratio of $e^{1.95} = 7.03$, and the association between 24 Month and Section 8 certificates (4.08) yielded an estimated odds ratio of $e^{3.33} = 27.94$, suggesting that individuals with Section 8 certificates were less likely to be in community housing as opposed to street/shelter housing at 6 Month, 12 Month and 24 Months.

Some may be interested in comparing community housing and independent housing. The baseline-category parameters of the $K-1$ equations in the model can be used to represent logit defined for pairs of different response categories as follows

$$\log\frac{P(\text{Community})}{P(\text{Independent})} = \log\frac{P(\text{Community})}{P(\text{Street/Shelter})} - \log\frac{P(\text{Independent})}{P(\text{Street/Shelter})}.$$

Thus,

$$\log\frac{\hat{P}(\text{Community})}{\hat{P}(\text{Independent})} = 1.40 + 1.07\text{Time}_1 + 0.12\text{Time}_2 - 0.35\text{Time}_3 + 0.65\text{Section 8} \\ - 4.59\text{Time}_1*\text{Section 8} - 5.47\text{Time}_2*\text{Section 8} - 4.62\text{Time}_3*\text{Section 8} + \hat{b}_{itK}.$$

The posterior means for diagonal element matrices of $\hat{A}_i$ were given by

$$\hat{A}_{it} = \begin{cases} \text{diag}(29.96, 0.85), & \text{Control group,} \\ \text{diag}(1.40, 3.22), & \text{Section 8.} \end{cases}$$

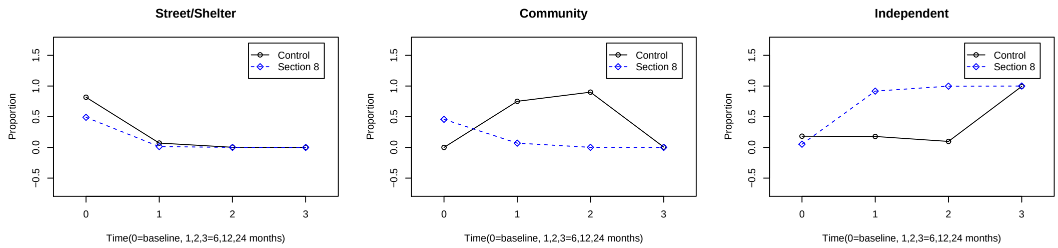

Figure 2: *Model fit of marginal proportions under two group (Section 8 and control) for response 1 (street/shelter), 2 (community) and 3 (independent housings), respectively.*

In the GARMs, the posterior mean for Ψ_{itj} is given by

$$\hat{\Psi}_{itj} = \begin{cases} \begin{pmatrix} 0.55 & 0.02 \\ -0.15 & 0.17 \end{pmatrix}, & \text{for Control group,} \\ \begin{pmatrix} 0.56 & 3.16 \\ 2.00 & -0.78 \end{pmatrix}, & \text{for Section 8 group.} \end{cases}$$

Figure 2 compares fitted marginal probabilities for Section 8 group versus control group. In the street/shelter housing, two estimated marginal probabilities decreased as the month increased. In the community housing and independent housing, there were many difference between Section 8 and control groups.

5. Conclusion

We proposed Bayesian baseline-category logit random effects models for longitudinal nominal data. In the models, the modified Cholesky decomposition (MCD) was used to decompose the random effects covariance matrix to the generalized autoregressive matrices (GARMs) and innovation variance matrices (IVMs). The GARMs account for serial correlations of nominal outcomes, and the IVMs explain prediction error variances. The MCD represents a computationally attractive approach and provides a better fit than the competing random intercept model with a homogeneous covariance. The proposed models also were fitted using a Bayesian approach. McKinney homeless research project (MHRP) data were analyzed using our proposed models. We fitted five baseline-category logit models to compare. Among the models, the model with a heteroscedastic AR(1) random effects covariance matrix was the best fit to our data. The estimated conditional probabilities for three groups were different trends as months increased.

Acknowledgments

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-2019R1F1 A1058553 for Keunbaik Lee).

References

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

Chen B, Yi GY, and Cook RJ (2009). Likelihood analysis of joint marginal and conditional models for longitudinal categorical data, *Canadian Journal of Statistics*, **37**, 182–205.

Daniels MJ and Gatsonis C (1997). Hierarchical polytomous regression models with applications to health services research, *Statistics in Medicine*, **16**, 2311–2325.

Daniels JM and Zhao YD (2003). Modeling the random effects covariance matrix in longitudinal data, *Statistics in Medicine*, **22**, 1631–1647.

Hartzel J, Agresti A, and Caffo B (2001). Multinomial logit random effects models, *Statistical Modelling*, **1**, 81–102.

Hedeker D (2003). A mixed-effects multinomial logistic regression model, *Statistics in Medicine*, **22**, 1433–1446.

Hedeker D and Gibbons RD (2006). *Longitudinal Data Analysis*, Wiley, Hoboken, New Jersey.

Hurlburt MS, Wood PA, and Hough RL (1996). Providing independent housing for the homeless mentally ill: A novel approach to evaluating long-term longitudinal housing patterns, *Journal of Community Psychology*, **24**, 291–310.

Lee K (2013). Bayesian modeling of random effects covariance matrix for generalized linear mixed models, *Communication for Statistical Applications and Methods*, **20**, 235–240.

Lee K, Cho H, Kwak MS,and Jang EJ (2020). Estimation of covariance matrix of multivariate longitudinal data using modified Choleksky and hypersphere decompositions, *Biometrics*, **76**, 75–86.

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

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

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

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 generalized linear models for multivariate normal covariance matrix, *Biometrika*, **87**, 425–435.

Revelt D and Train K (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, **80**, 647–657.

Theil H (1969). A multinomial extension of the linear logit model, *International Economic Review*, **10**, 251–259.

Theil H (1970). On the estimation of relationships involving qualitative variables, *American Journal of Sociology*, **76**, 103–154.

Received September 19, 2019; Revised December 2, 2019; Accepted December 24, 2019