ARMA Cholesky Decompositions for Random Effects Covariance Matrix in Marginalized Random Effects Models

Dasom Kang
Thesis advisor: Dr. Keunbaik Lee

Department of Statistics, Sungkyunkwan University

June 09, 2017

OUTLINE

- **Introduction**
- **Proposed Model**
- **Real data analysis**
- **Conclusion**

Introduction

Longitudinal study

A longitudinal study refers to repeated measures from same subjects observed over time.

- Controlling serial dependence is important part of longitudinal study.
- The within-subject measurements are typically not independent.

INTRODUCTION

Marginalized Random effect Models(MREMs)

The **MREMs** are commonly used to analyze longitudinal categorical data when the population-averaged effects is of interest. (Heagerty, 1999)

- Random effects capture both the correlation between the responses.
- Likelihood-based approach.
- Issues of modeling of random effects covariance matrix.

Introduction

Problems of random effects covariance matrix

Modelling a covariance matrix Σ_i is difficult because of

(a) the strong possibility of high-dimensionality

(b) the constraint that Σ_i must be positive definite.

- Several methods are proposed to solve the problems.
  - Modified Cholesky decomposition **(MCD)**, (Pourahmadi, 1999)
  - ARMA Cholesky decomposition **(ARMACD)**, (Lee et al.,2017)

Introduction

ARMA Cholesky decomposition

- The models provide for a wide variety of structures in the covariance but can be specified using small number of parameters. (Rochen and Helms, 1989)
- Parsimony in parameterization makes the model interpretation easy and provides stable estimation of parameters. (Lee et al., 2016)

Proposed Model

Proposed Model

Marginalized random effects models for longitudinal binary data using ARMA Cholesky Decomposition

Notation

- Y_{it} : the response for subject i (i =1, ..., N) at time t (t =1, ..., n_i).
- β : $p \times 1$ coefficients vector of x_{it}
- b_{it} : the random effects

Proposed Model

Marginalized Random effect models (MREMs)

$$
\begin{aligned}
\text{Marginal mean model} :& g(\mu_{it}^{M}) = x_{it}^{T}\beta, \\
\text{Dependence model} :& g(\mu_{it}^{c}(b_{it})) = \Delta_{it} + b_{it}, \\
& b_i \sim N(\mathbf{0}, \Sigma_i),
\end{aligned}
$$

where $\mu_{it}^{M} = E(Y_{it}|x_{it}),\ \mu_{it}^{c}(b_{it}) = E(Y_{it}|x_{it}, b_{it}),$

g($\cdot$) : link function.

PROPOSED MODEL

The parameter $\varDelta_{it}$

- Δ_{it} $\sim$ **function(** β, σ_{itt} **)**

$$
\begin{aligned}
E(Y_{it}|x_{it}) &= E(E(Y_{it}|x_{it}, b_{it})),\\
\Leftrightarrow \quad P(Y_{it}=1|x_{it}) &= E[P(Y_{it}=1|x_{it}, b_{it})],\\
&= \int P(Y_{it}=1|x_{it}, b_{it}) f(b_{it}) db_{it},\\
\Leftrightarrow \qquad \frac{e^{x_{it}^T\beta}}{1+e^{x_{it}^T\beta}} &= \int \frac{e^{\varDelta_{it}+b_{it}}}{1+e^{\varDelta_{it}+b_{it}}} f(b_{it}) db_{it},
\end{aligned}
$$

where $f(b_{it}) \sim N(\mathbf{0}, \sigma_{itt})$.

Proposed Model

- $\mathrm{P}_{it}^{M} = \int P_{it}^{c}(b_{it}) f(b_{it}) db_{it}$.
- Let $h(\Delta_{it}) = \int P_{it}^{c}(b_{it}) f(b_{it}) db_{it} - P_{it}^{M}$.
- Estimates of Δ_{it} can be obtained using Newton-Raphson algorithm as follows,

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

where $\dfrac{\partial h(\Delta_{it}^{(j)})}{\partial \Delta_{it}^{(j)}} = \displaystyle\int p_{it}^{c}(b_{it})(1 - p_{it}^{c}(b_{it})) f(b_{it}) db_{it}$.

Proposed Model

ARMA Cholesky Decomposition

- To model Σ_i, we assume that,

$$b_{i1} = e_{i1},$$
$$b_{it} = \sum_{j=1}^{t-1} \phi_{itj} b_{ij} + \sum_{j=1}^{t-1} l_{itj} e_{ij} + e_{it}, \quad \text{for } t = 2, \ldots, n_i \tag{1}$$

where $\mathrm{e}_i = (e_{i1}, ..., e_{in_i})^T \sim N(\mathbf{0}, D_i)$ *with*
$\mathrm{D}_i = diag(\sigma^2_{i1}, ..., \sigma^2_{in_i})$.

Proposed Model

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\phi_{i21} & 1 & \cdots & 0 \\ -\phi_{i31} & -\phi_{i32} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\phi_{in_i1} & -\phi_{in_i2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} b_{i1} \\ \\ \vdots \\ \\ b_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ l_{i21} & 1 & \cdots & 0 \\ l_{i31} & l_{i32} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{in_i1} & l_{in_i2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} e_{i1} \\ \\ \vdots \\ \\ e_{in_i} \end{pmatrix}$$

$$T_i b_i = L_i e_i \tag{2}$$

From (2), we have

$$\text{var}(\text{T}_i b_i) = var(L_i e_i) <=> T_i \Sigma_i T_i^T = L_i D_i L_i^T$$

$$\Sigma_i = T_i^{-1} L_i D_i L_i^T T_i^{-T}. \tag{3}$$

Proposed Model

The generalized autoregressive parameters(GARPs), generalized moving average parameters(GMAPs) and the innovation variances(IVs) can be modeled using time and/or subject-specific covariate vectors w_{itj}, z_{itj} and h_{it} by setting

$$\phi_{itj} = w_{itj}^T \alpha, \quad l_{itj} = z_{itj}^T \gamma, \quad \log \sigma_{it}^2 = h_{it}^T \lambda. \tag{4}$$

Maximum Likelihood Estimation

- $\theta = (\beta, \alpha, \gamma, \lambda)^T$.
- We use quasi-Newton methods to solve the likelihood equations.

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

where

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

Maximum Likelihood Estimation

Detailed calculations of the log-likelihood function

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

$$\frac{\partial logL(\theta;y)}{\partial\alpha_k} = \sum_{i=1}^{N} L^{-1}(\theta;y_i)\int L(\theta,b_i;y_i)\left\{\sum_{t=1}^{n_i}(y_{it}-P_{it}^c(b_{it}))\frac{\partial\Delta_{it}}{\partial\alpha_k} - \sum_{t=1}^{n_i}\frac{e_{it}}{\sigma_{it}^2}\frac{\partial e_{it}}{\partial\alpha_k}\right\}f(b_i)db_i,$$

$$\frac{\partial logL(\theta;y)}{\partial\gamma_k} = \sum_{i=1}^{N} L^{-1}(\theta;y_i)\int L(\theta,b_i;y_i)\left\{\sum_{t=1}^{n_i}(y_{it}-P_{it}^c(b_{it}))\frac{\partial\Delta_{it}}{\partial\gamma_k} - \sum_{t=1}^{n_i}\frac{e_{it}}{\sigma_{it}^2}\frac{\partial e_{it}}{\partial\gamma_k}\right\}f(b_i)db_i,$$

$$\frac{\partial logL(\theta;y)}{\partial\lambda_k} = \sum_{i=1}^{N} L^{-1}(\theta;y_i)\int L(\theta,b_i;y_i)\left\{\sum_{t=1}^{n_i}(y_{it}-P_{it}^c(b_{it}))\frac{\partial\Delta_{it}}{\partial\lambda_k} + \frac{1}{2}\sum_{t=1}^{n_i}\left(\frac{e_{it}^2}{\sigma_{it}^2}-1\right)h_{itk}\right\}f(b_i)db_i.$$

Maximum Likelihood Estimation

Derivatives of Δ_{it} with respect to $\beta, \alpha, \gamma, \lambda$

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

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

$$\Rightarrow \quad \frac{\partial \Delta_{it}}{\partial \beta} = \frac{P_{it}^{M}(1 - P_{it}^{M}) x_{it}}{\int P_{it}^{c}(b_{it})(1 - P_{it}^{c}(b_{it})) f(b_{it}) db_{it}},$$

$$\frac{\partial \Delta_{it}}{\partial \alpha_k} = \frac{\int P_{it}^{c}(b_{it}) \sum_{t=1}^{n_i} \frac{\partial e_{it}}{\partial \sigma_{it}^2} \frac{\partial e_{it}}{\partial \alpha_k} f(b_{it}) db_{it}}{\int P_{it}^{c}(b_{it})(1 - P_{it}^{c}(b_{it})) f(b_{it}) db_{it}},$$

$$\frac{\partial \Delta_{it}}{\partial \gamma_k} = \frac{\int P_{it}^{c}(b_{it}) \sum_{t=1}^{n_i} \frac{\partial e_{it}}{\partial \sigma_{it}^2} \frac{\partial e_{it}}{\partial \gamma_k} f(b_{it}) db_{it}}{\int P_{it}^{c}(b_{it})(1 - P_{it}^{c}(b_{it})) f(b_{it}) db_{it}},$$

$$\frac{\partial \Delta_{it}}{\partial \lambda_k} = \frac{\int P_{it}^{c}(b_{it}) \sum_{t=1}^{n_i} \frac{1}{2} \left(\frac{e_{it}^2}{\sigma_{it}^2} - 1 \right) h_{itk}}{\int P_{it}^{c}(b_{it})(1 - P_{it}^{c}(b_{it})) f(b_{it}) db_{it}}$$

Maximum Likelihood Estimation

Quasi-Monte carlo Integration

- Methods to approximate integral inside derivatives.

$$\int f(y_{it} \mid b_{it}) g(b_i) db_i \approx \frac{1}{M} f(y_{it} \mid b_{it}).$$

- We use **rnorm.sobol()** in the library fOptions (Wuertz, 2005)

Lung cancer data

Lung Cancer Study (Kim et al, 2012)

- 96 Patients (48 per arm).
- 2 Treatment arms: GEFITINIB and ERLOTINIB.
- Response rate: 47.9% (GEFITINIB), 39.6% (ERLOTINIB).
- Appetite was originally scored on a 4-point scale ranging from *I did not lose my appetite*(1) to *I lost my appetite*(4) .
- **We dichotomize the score between scores 1-2(0) 3-4(1).**
- Visit number. (TIME=0.0,0.1,...,0.9)

- Main Goal: To exam whether or not there was a negative impact of treatments on patients' Quality Of Life.

MODELS

- Description of 9 models based on ϕ_{itj}, α_{itj}, ℓ_{itj} and σ^2_{it}.

Model	GARP	GMAP	IV
AR(1)-1	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$		$log\sigma^2_{it} = \lambda_0$
AR(1)-2	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$		$log\sigma^2_{it} = \lambda_0 + \lambda_1 Grp_i$
AR(1)-3	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$		$log\sigma^2_{it} = \lambda_0 + \lambda_1 Time_{it}$
AR(2)-1	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)} + \alpha_1 I_{(\|t-j\|=2)}$		$log\sigma^2_{it} = \lambda_0$
ARMA(1,1)-1	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$	$l_{itj} = \gamma_0 I_{(\|t-j\|=1)}$	$log\sigma^2_{it} = \lambda_0$
ARMA(1,1)-2	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$	$l_{itj} = \gamma_0 I_{(\|t-j\|=1)}$	$log\sigma^2_{it} = \lambda_0 + \lambda_1 Grp_i$
ARMA(1,1)-3	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$	$l_{itj} = \gamma_0 I_{(\|t-j\|=1)}$	$log\sigma^2_{it} = \lambda_0 + \lambda_1 Time_{it}$
ARMA(1,1)-4	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)}$	$l_{itj} = \gamma_0 I_{(\|t-j\|=1)}$	$log\sigma^2_{it} = \lambda_0 + \lambda_1 Time_{it} + \lambda_2 Time^2_{it}$
ARMA(2,1)-1	$\phi_{itj} = \alpha_0 I_{(\|t-j\|=1)} + \alpha_1 I_{(\|t-j\|=2)}$	$l_{itj} = \gamma_0 I_{(\|t-j\|=1)}$	$log\sigma^2_{it} = \lambda_0 + \lambda_1 Time_{it}$

- Use maximized log likelihood values (MLE) and Akaike Information Criterion (AIC) to compare the models.

Model Fit

Table: log likelihoods and AICs for the models

	MREMs		GLMMs	
Model	Max. log likelihood	AIC	Max. log likelihood	AIC
AR(1)-1	-239.092	490.18	-245.569	503.13
AR(1)-2	-238.703	491.40	-244.458	502.69
AR(1)-3	-235.315	484.63	-245.325	504.65
AR(2)-1	-348.444	490.88	-243.413	500.82
ARMA(1,1)-1	-237.360	488.72	-241.455	496.91
ARMA(1,1)-2	-237.223	490.44	-241.141	498.28
ARMA(1,1)-3	-235.229	486.45	-240.602	497.20
ARMA(1,1)-4	-234.946	487.60	-240.602	497.20
ARMA(2,1)-1	-234.802	487.89		

Model Fit

- AICs in all models in MREMs were smaller than the AICs in GLMMs.
- MREMs worked well in lung cancer data analysis than GLMMs.

Table: Likelihood Ratio Test - ARMA(1,1)-3

	vs AR(1)-3	vs ARMA(1,1)-1	vs ARMA(1,1)-4
p-value	0.678	0.038	0.451

Model Fit Estimates

	ARMA(1,1)-3	AR(1)-3
Parameter: β		
Intercept	-1.213*(0.289)	-1.210*(0.289)
Time	-0.366(0.575)	-0.329(0.521)
Group	0.211(0.395)	0.210(0.394)
Time$\times$Group	-0.601(0.870)	-0.646(1.290)
GARPs or GMAPs :		
α_0	0.975*(0.152)	0.990*(0.081)
γ_0	-0.207(0.692)	
IVs parameters: λ		
Intercept λ_0	2.257(2.456)	0.173(0.651)
Time λ_1	-5.641*(2.442)	-7.567*(3.698)

*Significance at 95% confidence level.

Model Fit Results

- MREMs with random effects covariance matrix with a AR(1) depending on time was best fit among all models.
- For the fixed effects, only Intercept was significant under 5% significant level.
- Time, Group, and Interaction between Time and Group effect were not statistically significant.
- The covariance parameters of GARP and IV were significant. (α_0=0.990, S.E=0.081,λ_1= -7.567, S.E=3.698).
- It indicates that the random effects covariance matrix depended on time linearly.

Conclusion

- We have proposed MREMs for longitudinal binary data with the random effects covariance matrix. The random effects covariance matrix was factored by ARMA Cholesky decomposition into GARPs, GMAPs, and IVs.
- If IVs are positive, the random effects covariance matrix is positive definite. The number of parameters in the covariance matrix is reduced by considering linear models for GARPs, GMAPs and IVs.
- Using a quasi-Newton algorithm, estimation of parameters in the proposed models were calculated. Calculation of the integrations in likelihood and derivatives was conducted by quasi-Monte carlo integration.
- In analysis of lung cancer data, MREMs with random effects covariance matrix with a AR(1) depending on time was best fit among all models.

References

- Liang K. Y. and Zeger S. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.
- Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. (2002), Analysis of Longitudinal Data, *Oxford University Press. 2nd Edition*.
- Niederreiter, Harald. (1992), Random Number Generation and Quasi-Monte Carlo Methods, *Siam, Philadelphia, Pennsylvania*.
- Breslow, N. E. and Clayton, D. G. (1993), Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 125–134.
- Heagerty, P. J. (1999), Marginally specified logistic-normal models for longitudinal binary data, *Biometrics*, **55**, 688–698.
- Kim, S. T., Uhm, J. E., Lee, J., Sun, J. M., Sohn, I., Kim, S. W., Jung, S. H., Park, Y. H., Ahn, J. S., Park, K. and Ahn, M. J. (2012), Randomized phase II study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy, *Lung Cancer*, **75**, 82–88.

- Lee, K. and Sung S. (2014), Autoregressive Cholesky factor model for marginalized random effects model, *Communications for Statistical Applications and Methods*, **21**, 169–181.
- Pan, J. and MacKenzie, G. (2006), Regression models for covariance structure 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 generalised linear models for multivariate normal covariance matrix, *Biometrika*, **87**, 425–435.
- Zhang, W. and Leng, C. (2012), A moving average Cholesky factor model in covariance modelling for longitudinal data, *Biometrika*, **99**, 141–150.
- Lee, K., Baek, C., and Daniels, M. J. (2017), ARMA Cholesky factor models for the covariance matrix of linear models, *Computational Statistics & Data Analysis*. In press.
- Lee, K., Jung H., and Yoo, J. (2017), Modeling of the ARMA random effects covariance matrix in logstic random effects models, *working paper*.

- Lee, K. and Yoo, J. (2014), Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models, *Computational Statistics & Data Analysis*, **80**, 111–116.

- Kim, S. T., Uhm, J. E., Lee, J., Sun, J. M., Sohn, I., Kim, S. W., Jung, S. H., Park, Y. H., Ahn, J. S., Park, K. and Ahn, M. J. (2012), Randomized phase II study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy, *Lung Cancer*, **75**, 82–88.

- 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, J. (2014), Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models, *Computational Statistics & Data Analysis*, **80**, 111–116.

- Pan, J. and MacKenzie, G. (2006), Regression models for covariance structure 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.

Thank you.