

Covariance Structures for Multiply Repeated Measures Models

Hongmei Han¹, Jeff Burton¹, Robbie Beyl¹, Lei Zhang²
William D. Johnson¹

¹Pennington Biomedical Research Center, 6400 Perkins Rd, Baton Rouge, LA 70808

²University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216

Abstract

Consider a statistical model with a single outcome that is observed repeatedly under different circumstances. In a study employing an oral glucose tolerance test, for example, serum glucose and insulin levels may be measured just prior to ingesting a dose of glucose and then repeatedly every 30 minutes for a total of 120 minutes. The aim is to determine the pattern of change in glucose levels as an indication of how efficiently the individual is disposing the glucose from the blood. If an intervention is given and the process is repeated under this new condition, the data may be analyzed using a doubly repeated measures model. The use of structured patterns in the underlying covariance matrix for correlated residuals in statistical models involving repeated measures is well documented. In models involving multiple repeated measures, however, use of different covariance patterns for different conditions has not been fully discussed. The purpose of this paper is to fill that need by presenting an illustrative overview in terms of practical examples.

Key Words: Covariance patterns, General linear mixed models, Longitudinal data

1. Introduction

Repeated measures refer to response outcomes measured on the same experimental unit over time or under multiple conditions. The repeated measure factors can be randomized, or fixed, such as time or space. Repeated measures designs often enable more efficient estimates of residual variability due to a reduction in the number of parameters in the error variance, and thus more powerful test statistics. Such designs are widely encountered in behavioral science, agriculture, and biology. Despite advantages, the analyses of repeated measures are often complicated because there are a large number of possible correlation structures among repeated observations made on each experimental unit.

Doubly repeated measures are commonly used in research experiments. Comparing to single repeated measures, doubly repeated measures have two within-subject effects and thus the analysis may become more complex. Mixed models provide a flexible approach for analyzing such data by using a mixture of fixed and random effects and permitting a variety of covariance structures with the possibility of missing data and unequally spaced assessment times.

A single-blind randomized controlled trial was conducted to evaluate the effect of intervention on weight loss and insulin tolerance. Forty five participants were randomized into two groups, intervention (n=28) and control (n=17). Participants in intervention group had reduced food intake aiming to lose 10% initial body weight in the first 6 months. They were required to maintain their weight after the initial weight loss till the study ends at 24 months. Glucose tolerance were measured by a 75g 2-hour Oral Glucose Tolerance Test (OGTT) at baseline, month 12, and month 24. All tests were conducted in the morning following an overnight fast. After obtaining a fasting blood sample (Time 0), participants drank the 75g glucose solution (within 10 minutes) and blood sample were collected every 30 minutes for a total of 120 minutes for measurement of glucose and insulin. In this study, there is one between-subject factor, the treatment, and two within-subject factors: the time of clinical visit and the time within each visit (Time 0, +30, +60, +90, and +120). Changes from baseline on insulin were analyzed as the response variable. The log transformed insulin values were derived to have a normal distribution. Missing observations were considered to be missing completely at random.

There have been many publications addressing how to choose the right covariance structure for the single repeated data (Moser and Macchiavelli 2002; Kincaid, 2005; Littell et al. 2006), but few on how to properly model the data from doubly repeated design (Moser, 2004; Han and Johnson, 2012). In our previous work, we illustrated on how to select the best model with a doubly repeated design using PROC MIXED. This paper presents the application of both PROC MIXED and PROC GLIMMIX procedures to the analysis of a doubly repeated measure in a glucose tolerance study example. Both procedures allow users to choose from many different covariance structures. This paper studies proper use of the RANDOM and REPEATED statements with GROUP option in SAS to illustrate the analytical details on modeling doubly repeated measure. The use of different covariance patterns for different conditions is discussed with PROC GLIMMIX.

2. Methods

The typical linear mixed model notation is;

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\alpha}$ denotes fixed effects with design matrix \mathbf{X} , $\boldsymbol{\beta}$ random effects with $\boldsymbol{\beta} \sim N(0, \mathbf{G})$, and $\boldsymbol{\varepsilon}$ random error ($\text{var}(\boldsymbol{\varepsilon}) = \mathbf{R}$). So the variance of y , is $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$. The random portion of the model is fit by specifying the terms that define the random design matrix \mathbf{Z} and the structure of covariance matrices \mathbf{G} , \mathbf{R} , or both. In the MIXED procedure, RANDOM statement affects the \mathbf{G} matrix and REPEATED specifies the \mathbf{R} matrix.

In the generalized linear mixed model, the notation is;

$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}$ and $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{R}$. Therefore, the variance of y , is $\mathbf{V} = \text{var}(\boldsymbol{\varepsilon}) + \mathbf{R} \approx \mathbf{BZGZ}'\mathbf{B} + \mathbf{R}$, \mathbf{B} being a diagonal matrix of variance terms. In the GLIMMIX procedure, SAS code only contains a RANDOM statement which defines the \mathbf{Z} matrix and the structure of \mathbf{G} matrix. One can specify the `_RESIDUAL_` option to indicate the R-side random component that defined the \mathbf{R} matrix. Multiple RANDOM statements are possible in GLIMMIX.

2.1 Covariance Structures

Repeated measures have measurements taken on the same experimental unit the data are correlated with each other. The variance covariance structure could be modeled in \mathbf{G} or \mathbf{R} or both. SAS permits many covariance structures be modeled in PROC MIXED and

PROC GLIMMIX, such as CS, AR(1), UN, TOEP, et al. Three commonly used variance and covariance structure (CS, UN, and AR(1)) are shown below:

Unstructured (UN): The most complex one is unstructured which allows every term to be different. The parameter to fit such structure is $k(k+1)/2$.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \sigma_{12} & \dots & \sigma_{2k} \\ \sigma_{13} & \sigma_{12} & \sigma_3^2 & \dots & \sigma_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \sigma_{3k} & \dots & \sigma_k^2 \end{bmatrix}$$

Compound Symmetry (CS): The variances in CS are homogeneous and the correlation between any two measurements is constant.

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}$$

Autoregressive (1) AR(1): The variances in AR(1) are homogeneous, and correlations decline exponentially over time. Measurements made closer are more correlated than measurements made farther apart.

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{k-1} \\ \rho & 1 & \rho & \dots & \rho^{k-2} \\ \rho^2 & \rho & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{k-1} & \rho^{k-2} & \vdots & \rho & 1 \end{bmatrix}$$

In repeated measures, there is a tendency for measurements made close together across time to be more highly correlated than measurements made farther apart. Covariance structures that can accommodate changes in correlation over time, such as first-order autoregressive (AR(1)), heterogeneous variance autoregressive (ARH(1)) and antedependence structures (TOEPH) (Little et al 2006), may be more appropriate than CS.

2.1 Direct Product Covariance Structure

If A is an m -by- n matrix and B is a p -by- q matrix, then the Kronecker product $A \otimes B$ is the mp -by- nq block matrix.

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$

The example below shows the Kronecker product of two matrixes, a 2 by 2 unstructured matrix and a 2 by 2 compound symmetrical matrix. The final product is a 4 by 4 block matrix.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} & \rho \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \\ \rho \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} & \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \end{bmatrix}$$

Below is another example of the Kronecker product of two matrixes, a 3 by 3 unstructured matrix and a 2 by 2 unstructured matrix. The final product is a 6 by 6 block matrix.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{11} & a_{13}b_{11} & a_{11}b_{12} & a_{12}b_{12} & a_{13}b_{12} \\ a_{21}b_{11} & a_{22}b_{11} & a_{23}b_{11} & a_{21}b_{12} & a_{22}b_{12} & a_{23}b_{12} \\ a_{31}b_{11} & a_{32}b_{11} & a_{33}b_{11} & a_{31}b_{12} & a_{32}b_{12} & a_{33}b_{12} \\ a_{11}b_{21} & a_{12}b_{21} & a_{13}b_{21} & a_{11}b_{22} & a_{12}b_{22} & a_{13}b_{22} \\ a_{21}b_{21} & a_{22}b_{21} & a_{23}b_{21} & a_{21}b_{22} & a_{22}b_{22} & a_{23}b_{22} \\ a_{31}b_{21} & a_{32}b_{21} & a_{33}b_{21} & a_{31}b_{22} & a_{32}b_{22} & a_{33}b_{22} \end{bmatrix}$$

3. Illustrative Applications

3.1 Analysis Example with the MIXED procedure

First, PROC MIXED has the option to enable direct product covariance structures for two repeated factors. Let's consider the SAS code:

```
PROC MIXED Data = One;
  Class Trt Visit Time Subject;
  Model CFB = Time|Visit|Trt / DDFM = KenwardRoger;
  Repeated Visit Time / Subject = Subject Type = UN@CS R Rcorr;
Run;
```

The analysis invoked using this code assumes an unstructured covariance matrix (UN) for the levels of visit and CS for the levels of time. The 'DDFM = KenwardRoger' specifies using the Kenward-Rogers method for estimating the denominator degrees of freedom for some of the relevant test statistics. This is an especially useful method when there are missing values for some of the data.

The estimated UN covariance matrix for factor time is given in Table 1a, which contains 15 variance covariance parameter estimates. The estimated correlation is 0.39 for CS structure on factor visit. The final Kronecker product is shown in Table 1b. The diagonals of Table 1b are the estimated variances of the errors associated with 5 time measurements for 2 follow up visits.

Table 1a: Estimated UN covariance matrix of response on the repeated factor time

Row	Col1	Col2	Col3	Col4	Col5
1	0.16				
2	0.07	0.22			
3	0.02	0.06	0.25		
4	0.00	0.07	0.16	0.45	
5	0.00	0.09	0.07	0.27	0.41

Table 1b: Estimated covariance matrix for a model using the direct product code UN@CS to indicate an unstructured covariance configuration for Time and Compound Symmetry for Visit: Row/Column 1-5 correspond to the measurements made across the 5

times in follow up visit 1, Row/Column 6-10 correspond to the measurements made across the 5 times in follow up visit 2.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
1	0.16									
2	0.07	0.22								
3	0.02	0.06	0.25							
4	0.00	0.07	0.16	0.45						
5	0.00	0.09	0.07	0.27	0.41					
6	0.06					0.16				
7	0.03	0.09				0.07	0.22			
8	0.01	0.02	0.10			0.02	0.06	0.25		
9	0.00	0.03	0.06	0.17		0.00	0.07	0.16	0.45	
10	0.00	0.03	0.03	0.11	0.16	0.00	0.09	0.07	0.27	0.41

PROC MIXED allows three direct product structures for two within-subject factors: UN*UN, UN*CS, and UN*AR(1). We can interchange the order of two factors in the REPEATED statement to select CS or AR(1) for repeated factor 1, or for repeated factor 2. Using three information criteria, Akaike's information criterion (AIC), AIC corrected (AICc), and Bayesian information criterion (BIC), the "best" model can be selected among different covariance structures (Moser and Macchiavelli 2002).

Alternatively, PROC MIXED permits specifying a single effect while using "GROUP" specification to model doubly repeated measures. The analysis resulting from the following SAS code assumes separate covariance structures for different visits, with the CS for various time measurements.

```
PROC MIXED Data = One;
  Class Trt Visit Time Subject;
  Model CFB = Time|Visit|Trt / DDFM = KenwardRoger;
  Repeated Time / Subject = Subject*Visit Group = Visit Type = CS R Rcorr;
Run;
```

The estimated covariance matrices for two visits are given in Table 2. The number of estimated covariance parameters is reduced to 4 in this model. Employing another approach, PROC MIXED will estimate the covariance structure using information pooled over two visits if we leave out the 'GROUP =' option. The analysis further assumes identical covariance structures among the visits if the 'GROUP = Visit' option is omitted. The resulting number of estimated covariance parameters will be halved. We can experiment with various covariance structure, like CS, AR(1), or SP(POW), in search of improvement in model fitness.

Table 2: Covariance estimates for a model using separated CS structures by two follow up visits.

Follow up	Covariance Parameter	Subject	Group	Estimate
Visit 1	Variance	visit*subjectid	Active	.24
Visit 2	CS	visit*subjectid	Active	.08
Visit 2	Variance	visit*subjectid	Control	.16

CS *visit*subjectid* *Control* .06

3.2 Analysis Example with the GLIMMIX procedure

Linear mixed models are a special case in the family of generalized linear mixed models, where the conditional distribution is normal and the link function is the identity function. The GLIMMIX procedure models all random components of the model through the RANDOM statement.

There is generally considerable overlap in the covariance structures available through the TYPE= option in the RANDOM statement in PROC GLIMMIX. However, the Kronecker-type structures, as we described earlier in the MIXED procedure, are currently not supported in the GLIMMIX procedure.

Let's consider the SAS code:

```
PROC GLIMMIX Data = One;
  Class Trt Visit Time Subject;
  Model CFB = Time|Visit|Trt / DDFM = KenwardRoger;
  Random visit time/ subject=subjectid type=cs g v=1 ;
Run;
```

The analysis invoked using this code above has 2 G-side Covance parameters and 1 R-side parameter estimated. The estimated V matrix for the first subject is listed as Table 3.

Table 3: Estimated covariance matrix for a model with CS configuration for Time and Visit using GLIMMIX procedure: Row/Column 1-5 correspond to the measurements made across the 5 times in follow up visit 1, Row/Column 6-10 correspond to the measurements made across the 5 times in follow up visit 2.

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
1	0.28									
2	0.08	0.28								
3	0.08	0.08	0.28							
4	0.08	0.08	0.08	0.28						
5	0.08	0.08	0.08	0.08	0.28					
6	0.08					0.28				
7	0.03	0.08				0.08	0.28			
8	0.03	0.03	0.08			0.08	0.08	0.28		
9	0.03	0.03	0.03	0.08		0.08	0.08	0.08	0.28	
10	0.03	0.03	0.03	0.03	0.08	0.08	0.08	0.08	0.08	0.28

We can define R-side structure by modifying the above SAS code slightly. This is easily achieved by adding "RESIDUAL" option, as shown in SAS code below. Now one additional R-side parameter is required comparing to SAS code above.

```
PROC GLIMMIX Data = One;
  Class Trt Visit Time Subject;
  Model CFB = Time|Visit|Trt / DDFM = KenwardRoger;
  Random visit / subject=subjectid type=CS RESIDUAL g v=1 ;
  Random time / subject=subjectid type=CS g v=1 ;
```

Run;

Furthermore, we tried to fit the different covariance structure for each treatment group in GLIMMIX procedure, for exploration purpose. An indicate variable of group assignment (group 1 vs group 2) is required, as shown in the SAS code below. Unless the data clearly shows the need to fit different covariance structures by each group, it's not recommended to fit in such complex model.

```
PROC GLIMMIX data = One;
  Class Trt Visit Time Subject;
  Model CFB = Time|Visit|Trt / DDFM = KenwardRoger;
  Random visit*g1 / subject=subjectid type=CS;
  Random visit*g2 / subject=subjectid type=AR(1);
  Random time / subject=subjectid type=CS g v=1 ;
Run;
```

3.3 Multiple Repeated Measures

Occasional, some research designs study with more than 2 repeated factors. In order to simplify the modelling process, one can start with the data reduction, such as calculating the area under the curve for the 3rd repeated factor. After that, PROC MIXED and PROC GLIMMIX can be used to analyses the data as we described earlier from designs with two repeated factors.

3.4 Selecting the Covariance Structure

There are a few considerations when selecting the covariance structure. Rules of thumb are by parsimony, by meaning, and by Information Criteria (Kincard, 2005). Often time, complex model with too many parameters doesn't gain much efficiency and makes it less generalizable. In practice, it's not recommended to fit all possible structures and rely on the information criteria to pick the best one for you. Instead, with the consideration of the study design, treatment structures, and the meaning of the covariance structures, a few candidate structures can be safely decided. Information criteria can be further compared but comparable results from different structures should be expected.

4 Concluding Remarks

Much analytical efficiency can be gained by taking advantage of underlying covariance matrix in repeated measures. This paper discussed some model fitting strategies and the unique covariance structures in the example of a glucose tolerance study. PROC MIXED enables doubly repeated measures analysis with direct product covariance structures. Both MIXED and GLIMMIX can specify a single effect while using group statement to handle doubly repeated measures.

This paper has not dealt with the many diagnostic analyses that should be incorporated into the use of the repeated measures analyses. Residuals could be examined for the bell shaped distribution. The effects of influential observations and outliers on results should be examined as well. Further work will be done with simulation study in the future.

Acknowledgements

Supported by 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health which funds the Louisiana Clinical and Translational Science Center

References

- Han, H. and Johnson, W.D. (2012) Application of Doubly Repeated Measures Analyses in Drug Trials on Food Intake in JSM Proceedings, Biopharmaceutical Section. American Statistical Association. 830-840
- Kincaid, C. (2005) "Guidelines for selecting the covariance structure in mixed model analysis" SUGI 30 Proceedings: Paper 198-30. Available at: <http://www2.sas.com/proceedings/sugi30/198-30.pdf>
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger O. (2006) SAS for Mixed Models second edition. Cary, NC; SAS Institute Inc.
- Moser, E.B. (2004) Repeated measures modeling with PROC MIXED. Proceedings of the 29th SAS Users Group International Conference. Montreal, Canada, May 9-12, 2004. Available at: <http://www2.sas.com/proceedings/sugi29/188-29.pdf>
- Moser, E. B., and Macchiavelli, R. E. (2002), "Model selection techniques for repeated measures covariance structures," Proceedings of Fourteenth Annual Kansas State University Conference on Applied Statistics in Agriculture, 17-31.
- SAS OnlineDoc 9.1.3. <http://support.sas.com/onlinedoc/913/docMainpage.jsp>