Australian Journal of Basic and Applied Sciences, 2(4): 1228-1238, 2008
ISSN 1991-8178

The Impact of Restrected Our Analysis of Repeated Measures Design to the Two Stander Covariance Structures with and Without Missing Data

Ali Hussein AL-Marshadi

Department of Statistics, King Abdul Aziz University, Jeddah, Saudi Arabia.

Abstract: This article considers the analysis of experiment of repeated measures design that is used frequently in different area of research. We used simulation study to investigate a common problem that is more frequently researchers, assuming one of two correlated error model in analyzing repeated measures design. This study evaluate the impact of restrict our options of covariance structure to the two correlated error model options of the split-plot analysis approach and the unstructured analysis approach by assessing the size and power of the tests of the fixed effects. The simulation results show that overall, the right analysis approach provided the best power and in the same time it has a very good control of the nominal Type I error rate for all the effects in both cases of balanced and unbalanced data. The main result of our article is that overall the right analysis approach is more powerful than the unstructured analysis approach when both the approaches yield satisfactory Type I error control comparing to the split-plot analysis approach with the Kenward-Roger method.

Key words: Repeated Measures Design, Kenward-Roger method, Restricted maximum likelihood (REML)

INTRODUCTION

The suitable analysis of a study is very important factor that contribute to the success of the study under consideration. A poorly analyzed study can produce misleading or inaccurate results.

In repeated measures designs, many measurements are made on the same subject. Therefore they need special attention, since in practice the observations within each subject are likely to be correlated. The covariance structure of such observed data makes repeated measures design experiments data analysis different from other factorial experiments. More frequently researchers, assuming one of the two following correlated error structure with the mixed model:

- Independent errors model which assumes that all observations within a subject in a repeated measures design, are equally correlated (compound symmetry condition). This assumption is not justifiable in most typical repeated measures experiment. This approach is called split-plot analysis approach for analyzing the repeated measures design. Huynh and Feldt[8] showed that conditions required for the usual analysis of variance for repeated measures designs were less stringent than the compound symmetry condition. Type-H structure for the covariance matrix of the repeated measures factor is required to have valid F-test for repeated measures effect and the interactions involving the repeated measures effect with this approach[8]. Therefore, special methods of analysis are usually needed to accommodate the correlation structure of the repeated measures when this assumption is not satisfied[9].
- Unstructured correlated error model which is conceptually similar to the correlation structure assumed when we use the unstructured multivariate approach (MANOVA) to analyze repeated measures data because they are based on the same estimate of covariance structure. This approach is called unstructured analysis approach for analyzing the repeated measures design. This approach has the advantage over the split-plot approach in that it does not require any specific structure on the covariance matrix of the repeated measures factor and it has the advantage over the MANOVA in that it does not ignore cases where any of the repeated measurements are missing.

The MIXED procedure of the SAS System[13] is a standard tool for analyzing correlated data. The MIXED procedure of the SAS System allowed statisticians to utilize an inclusive array of correlated error models. The MIXED procedure of the SAS System has different options for modeling the covariance structure

Corresponding Author: Ali Hussein AL-Marshadi, Department of Statistics, King Abdul Aziz University, Jeddah, Saudi Arabia.
Email: almarshadi-a@hotmail.com

of the experiments of repeated measures design. The MIXED procedure of the SAS System can be used to develop either maximum likelihood (ML) or restricted maximum likelihood (REML) estimates in order to complete the analysis of the experiments of repeated measures design. REML estimation is generally preferred to ML and it is the default in The MIXED procedure. Therefore, it used to complete the analysis in the simulation study. Because lot of effort is usually needed to decide what the suitable covariance structure of the data is at the beginning of the statistical analysis, more frequently researchers assume one of the two common correlated error models the split-plot analysis approach and the unstructured analysis approach.

Our research objective is finding the impact of restrict our options of covariance structure to the two options of the split-plot analysis approach and the unstructured analysis approach that are built up to most "point and clack" statistical software. A simulation study was conducted to evaluate the impact of restrict our options of covariance structure to the two options of the split-plot analysis approach and the unstructured analysis approach by assessing the size and power of the tests of the fixed effects in both cases of balanced and unbalanced data.

In general form, the mixed effects linear model can be written as[7,12]:

$$Y = X\beta + ZU + e \tag{1.1}$$

where:

$\beta = p \times 1$ vector of fixed effects.
$U = q \times 1$ vector of random effects.
$e = n \times 1$ vector of residuals.
$X = n \times p$ design matrix for fixed effects.
$Z = n \times q$ design matrix for random effects.
U $\sim N(0,G)$, $N(0,R)$,
$Y \sim N(X\beta,V)$ $V = ZGZ' + R$.

when V is known, the best linear unbiased estimators (BLUE) of estimable functions $h`\beta$ of the fixed effects in (1.1) are given by

$$h`\hat{\beta} = h`(XV^{-1}X)^{-}XV^{-1}Y, \tag{1.2}$$

with $$\operatorname{var}(h`\hat{\beta}) = h`(X`V^{-1}X)^{-}h. \tag{1.3}$$

In most applications V is unknown. Therefore, It is estimated from the data where estimators based on (1.2) are not generally BLUE[6]. Various procedures were proposed for testing hypotheses on fixed effects in mixed models with unknown V, most of which assume that V is estimated by the REML method[2,4,11]. The resulting estimates of fixed effects are often referred to as empirical BLUE (eBLUE)[6]. Standard error estimates based on (1.3) are biased downwards when V replaced by its estimate[10]. Fixed effects are estimated

based on (1.2), with V replaced by a plug-in REML estimate. Null hypotheses of the form $H_0 : h`\beta = 0$ are tested by

$$F = \frac{\hat{\beta}`h[h`(X\hat{V}^{-1}X)^{-}h]^{-1}h`\hat{\beta}}{rank(h)} \,\square\, F_{(rank(h),\upsilon)} \tag{1.4}$$

when $rank(h) > 1$. In general, the test statistics in (1.4) only have approximate F-distribution. The approximate denominator degree of freedom υ of F-distribution can be determined using one of the four different methods implemented in MIXED procedure of SAS. The four methods of the approximations are residual method, containment method (this is the default in MIXED), extended Satterthwaite[14] method of Giesbrecht and Burns[4] and Fai and Cornelius[2], and Kenward-Roger method[11]. The Kenward-Roger method was considered in this paper for approximating the denominator degrees of freedom since Kenward and Roger found good performance of their method across a number of designs[11], and also, Guerin and Stroup recommended using the Kenward-Roger method as standard operating procedure[5].

MATERIALS AND METHODS

The following model reflects the basic experiment of repeated measures design which is a special case of the mixed model (1.1):

$$y_{ijk} = \mu + \alpha_i + d_{ij} + \tau_k + (\alpha\tau)_{ik} + \varepsilon_{ijk} \tag{2.1}$$

where $i = 1, 2, \ldots, t; j = 1, 2, \ldots, r; k = 1, 2, \ldots, a.$

μ is an overall mean parameter; α_i is the *ith* treatment main effect, d_{ij} is the random effect associated with the *jth* subject in the *ith* treatment main effect, assumed i.i.d $N(0, \sigma^2_{subj})$, τ_k is the *ith* time main effect, $(\alpha\tau)_{ik}$ is the interaction between treatments and time, and ε_{ijk} is within subjects errors associated with the *jth* subject in the *ith* treatment main effect at the *kth* time effect. Within subjects errors are potentially correlated. Therefore, the vector $\varepsilon_{..}$ Is defined as follow $\varepsilon_{ij} = \begin{bmatrix} \varepsilon_{ij1} & \varepsilon_{ij2} & \ldots & \varepsilon_{ija} \end{bmatrix} \approx MVN(\mathbf{0}, \Sigma)$ where Σ is the covariance matrix among ε_{ijk} within the *ijth* subject. The vectors ε_{ij} are assumed to be mutually independent. Also, *dij*, and ε_{ijk} are assumed to be independent. Alternatively, we can express model (2.1) as cell mean model:

$$y_{ijk} = \mu_{ik} + d_{ij} + \varepsilon_{ijk} \tag{2.2}$$

where, $i = 1, 2, \ldots, t; j = 1, 2, \ldots, n; k = 1, 2, \ldots, a;$ and $\mu_{ik} = \alpha_i + \tau_k + (\alpha\tau)_{ik}$ Therefore, the vector $\mathbf{y}_{ij}$ s defined as follow $\mathbf{y}_{ij} = \begin{bmatrix} y_{ij1} & y_{ij2} & \ldots & y_{ija} \end{bmatrix}' \approx MVN(\mu_i, \mathbf{J}\sigma^2_{subj} + \Sigma)$ where μ_i is the vector of means at *kth* time for the *ith* treatment, i.e. $\mu_i = \begin{bmatrix} \mu_{i1} & \mu_{i2} & \ldots & \mu_{ia} \end{bmatrix}$ A different covariance matrix structures of Σ will be used to simulate correlated error models for the simulated experiment. The following experimental design was consider because of its practical relevance. The design of the simulated experiment is described below:

The treatments were arranged in a basic form of repeated measures design which consists of a completely randomized experimental design with data collected in a sequence of equally spaced points in time. The design of the simulated experiment is described in diagram 1, which consists of

$t = 3$ treatments
n = (5 or 10) subjects per treatment level
$a = 7$ repeated measures within each treatment level.

Diagram 1. Illustration of the design of the simulated experiment

Treatments	Time						
A	1	2	3	4	5	6	7
B	1	2	3	4	5	6	7
C	1	2	3	4	5	6	7

The repeated measures analyses for this design can be implemented by the following example SAS code:

```
PROC MIXED DATA = one;
CLASS treatment  subject  time;
MODEL y = treatment time treatment * time / ddfm = kr;
REPEATED time / type = UN subject = subject (treatment);
```

An issue concerns the estimation of the variance-covariance components of G and R. In the repeated measures design experiments, the general form of $V(Y) = V = ZGZ' + R$ s block diagonal of $J\sigma^2_{subj} + \Sigma$, with each block corresponding to a subject. When covariance structure of both the unstructured analysis approach (Unstructured covariance structure) and the split-plot analysis approach (Compound Symmetry covariance structure) are considered for Σ in the analyses, the component of Σ contain the between subject variance (σ^2_{subj}), since in case of the unstructured analysis approach, the unstructured covariance structure is the most general one possible and in case of the split-plot analysis approach, the between subject variance (σ^2_{subj}) and the constant correlation between observation on the *ijth* subject are confound. In this case, model (2.1) reduces to

$$y_{ijk} = \mu + \alpha_i + \tau_k + (\alpha\tau)_{ik} + \varepsilon_{ijk} \tag{2.3}$$

where, $i = 1, 2, \ldots, t; j = 1, 2, \ldots, r; k = 1, 2, \ldots, a$. In other covariance structure, we will follow a typical strategy that suggested dropping d_{ij} from model (2.2) and use model (2.3) instead in order to avoid non-convergence problem that we experienced during our simulation. This action will give us the chance to assess the impact of dropping d_{ij} from the model when it is required by theory.

The Simulation Study:

A simulation study of PROC MIXED's mixed model analysis of repeated measures data was conducted to evaluate the impact of restrict our options of covariance structure to the two options of the split-plot analysis approach and the unstructured analysis approach in both cases of balanced and unbalanced data. Unbalancedness (Missing Data) was generated by randomly dropping certain number of observations. In our investigations, the evaluation of the analysis was in terms of control of Type I error, and the power of the tests of the fixed effects from all the analyses. Kenward-Roger method was used for computing the denominator degrees of freedom for the tests of fixed effects from all the analyses. In case of a valid alternative hypothesis, the values of fixed effect used under the alternative are summarized in Table 1. Also, the percentage of number of times that REML failing to converge with normal situation, where the PROC MIXED procedure used REML without any interfering, was reported.

Table 1: The main effect of treatment, the main effect of time, and the interaction effect of treatment and time that were used in the simulation study.

Levels of Treatment effect	Levels of Time Effect							Treatment Main effect
	1	2	3	4	5	6	7	
1	2	0	0	0	0	0	-2	0
2	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	-1	0
Time Main effect	1	0	0	0	0	0	-1	

Correlated multivariate normal data were generated according to model (2.1). There were 24 scenarios to generate data involving eight covariance structures with different settings of covariance matrix parameter values for each covariance structure and two different sample sizes (n = 5 and 10 subjects per treatment). The covariance structures were Compound Symmetry (CS), Heterogeneous Compound Symmetry (CSH), First-Order Autoregressive (AR(1)), Heterogeneous First-Order Autoregressive (ARH(1)), Independent Errors (VC), Banded Main Diagonal (UN(1)), Toeplitz (TOEP), and Unstructured (UN). The 12 settings of the covariance matrix are given in Table 2 which can be categorized to eight covariance structures. The first one, (Setting No. 1.1 and 1.2) represents Compound Symmetry (CS) covariance structure with different choices of parameters. The second one, (Setting No. 2.1 and 2.2) represents First-Order Autoregressive (AR(1)) covariance structure with different choices of parameters. The third one, (Setting No. 3.1 and 3.2) represents

Heterogeneous First-Order Autoregressive (ARH(1)) covariance structure with different choices of parameters. The fourth one, (Setting No. 4.1 and 4.2) represents Heterogeneous Compound Symmetry (CSH) covariance structure with different choices of parameters. The fifth one, (Setting No. 5.1) represents Independent Errors (VC) covariance structure. The sixth one, (Setting No. 6.1) represents Banded Main Diagonal (UN(1)) covariance structure. The seventh one, (Setting No. 7.1) represents Toeplitz (TOEP) covariance structure. The eighth one, (Setting No. 8.1) represents Unstructured (UN) covariance structure. For each scenario, we simulated 4000 datasets.

Table 2: The 12 settings of eight Covariance Matrix structures used in the Simulations

Setting No.	Covariance Matrix	Setting No.	Covariance Matrix
1.1	$\begin{bmatrix} 16 & 8 & 8 & 8 & 8 & 8 & 8 \\ 8 & 16 & 8 & 8 & 8 & 8 & 8 \\ 8 & 8 & 16 & 8 & 8 & 8 & 8 \\ 8 & 8 & 8 & 16 & 8 & 8 & 8 \\ 8 & 8 & 8 & 8 & 16 & 8 & 8 \\ 8 & 8 & 8 & 8 & 8 & 16 & 8 \\ 8 & 8 & 8 & 8 & 8 & 8 & 16 \end{bmatrix}$	1.2	$\begin{bmatrix} 16 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 16 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 16 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 16 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 16 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 16 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 16 \end{bmatrix}$
2.1	$\begin{bmatrix} 16 & 8 & 4 & 2 & 1 & 0.5 & 0.25 \\ 8 & 16 & 8 & 4 & 2 & 1 & 0.5 \\ 4 & 8 & 16 & 8 & 4 & 2 & 1 \\ 2 & 4 & 8 & 16 & 8 & 4 & 2 \\ 1 & 2 & 4 & 8 & 16 & 8 & 4 \\ 0.5 & 1 & 2 & 4 & 8 & 16 & 8 \\ 0.25 & 0.5 & 1 & 2 & 4 & 8 & 16 \end{bmatrix}$	2.2	$\begin{bmatrix} 16 & 12.8 & 10.24 & 8.192 & 6.5536 & 5.24288 & 4.194304 \\ 12.8 & 16 & 12.8 & 10.24 & 8.192 & 6.5536 & 5.24288 \\ 10.24 & 12.8 & 16 & 12.8 & 10.24 & 8.192 & 6.5536 \\ 8.192 & 10.24 & 12.8 & 16 & 12.8 & 10.24 & 8.192 \\ 6.5536 & 8.192 & 10.24 & 12.8 & 16 & 12.8 & 10.24 \\ 5.24288 & 6.5536 & 8.192 & 10.24 & 12.8 & 16 & 12.8 \\ 4.194304 & 5.24288 & 6.5536 & 8.192 & 10.24 & 12.8 & 16 \end{bmatrix}$
3.1	$\begin{bmatrix} 4 & 3 & 2 & 1.25 & 0.75 & 0.4375 & 0.25 \\ 3 & 9 & 6 & 3.75 & 2.25 & 1.3125 & 0.75 \\ 2 & 6 & 16 & 10 & 6 & 3.5 & 2 \\ 1.25 & 3.75 & 10 & 25 & 15 & 8.75 & 5 \\ 0.75 & 2.25 & 6 & 15 & 36 & 21 & 12 \\ 0.4375 & 1.3125 & 3.5 & 8.75 & 21 & 49 & 28 \\ 0.25 & 0.75 & 2 & 5 & 12 & 28 & 64 \end{bmatrix}$	3.2	$\begin{bmatrix} 4 & 4.8 & 5.12 & 5.12 & 4.9152 & 4.58752 & 4.194304 \\ 4.8 & 9 & 9.6 & 9.6 & 9.216 & 8.6016 & 7.86432 \\ 5.12 & 9.6 & 16 & 16 & 15.36 & 14.336 & 13.1072 \\ 5.12 & 9.6 & 16 & 25 & 24 & 22.4 & 20.48 \\ 4.9152 & 9.216 & 15.36 & 24 & 36 & 33.6 & 30.72 \\ 4.58752 & 8.6016 & 14.336 & 22.4 & 33.6 & 49 & 44.8 \\ 4.194304 & 7.86432 & 13.1072 & 20.48 & 30.72 & 44.8 & 64 \end{bmatrix}$
4.1	$\begin{bmatrix} 4 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 9 & 6 & 7.5 & 9 & 10.5 & 12 \\ 4 & 6 & 16 & 10 & 12 & 14 & 16 \\ 5 & 7.5 & 10 & 25 & 15 & 17.5 & 20 \\ 6 & 9 & 12 & 15 & 36 & 21 & 24 \\ 7 & 10.5 & 14 & 17.5 & 21 & 49 & 28 \\ 8 & 12 & 16 & 20 & 24 & 28 & 64 \end{bmatrix}$	4.2	$\begin{bmatrix} 4 & 4.8 & 6.4 & 8 & 9.6 & 11.2 & 12.8 \\ 4.8 & 9 & 9.6 & 12 & 14.4 & 16.8 & 19.2 \\ 6.4 & 9.6 & 16 & 16 & 19.2 & 22.4 & 25.6 \\ 8 & 12 & 16 & 25 & 24 & 28 & 32 \\ 9.6 & 14.4 & 19.2 & 24 & 36 & 33.6 & 38.4 \\ 11.2 & 16.8 & 22.4 & 28 & 33.6 & 49 & 44.8 \\ 12.8 & 19.2 & 25.6 & 32 & 38.4 & 44.8 & 64 \end{bmatrix}$
5.1	$\begin{bmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 16 \end{bmatrix}$	6.1	$\begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 49 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 64 \end{bmatrix}$
7.1	$\begin{bmatrix} 16 & 11.2 & 8 & 6.4 & 4.8 & 3.2 & 1.6 \\ 11.2 & 16 & 11.2 & 8 & 6.4 & 4.8 & 3.2 \\ 8 & 11.2 & 16 & 11.2 & 8 & 6.4 & 4.8 \\ 6.4 & 8 & 11.2 & 16 & 11.2 & 8 & 6.4 \\ 4.8 & 6.4 & 8 & 11.2 & 16 & 11.2 & 8 \\ 3.2 & 4.8 & 6.4 & 8 & 11.2 & 16 & 11.2 \\ 1.6 & 3.2 & 4.8 & 6.4 & 8 & 11.2 & 16 \end{bmatrix}$	8.1	$\begin{bmatrix} 4 & 2.4 & 4.8 & 8 & 8.4 & 7 & 4.96 \\ 2.4 & 9 & 2.4 & 1.5 & 2.7 & 7.35 & 10.8 \\ 4.8 & 2.4 & 16 & 3.4 & 10.08 & 15.4 & 6.48 \\ 8 & 1.5 & 3.4 & 25 & 18.9 & 16.45 & 9.2 \\ 8.4 & 2.7 & 10.08 & 18.9 & 36 & 4.62 & 22.56 \\ 7 & 7.35 & 15.4 & 16.45 & 4.62 & 49 & 16.24 \\ 4.96 & 10.8 & 6.48 & 9.2 & 22.56 & 16.24 & 64 \end{bmatrix}$

SAS (Version 9.1) PROC IML code was written to generate the datasets according to the described design.

We will consider a case when we have 3 subjects per treatment as an example to explain the process of generating the datasets. A 9 7x1 vector of standard normal random deviates were generated using SAS's NORMAL function. Denoted the vector:

$$\varepsilon_{ij} = \left[\varepsilon_{1ij}\, \varepsilon_{2ij}\, \varepsilon_{3ij}\, \varepsilon_{4ij}\, \varepsilon_{5ij}\, \varepsilon_{6ij}\, \varepsilon_{7ij} \right]'$$

Where $i = 1,2,3$; $j = 1,2,3$ Note that the 9 represents the 3 subjects for each of the 3 treatments and the 7 represents the 7 levels of time effect within each of the 3 levels of treatment effect. Also, a 9 7x1 vectors, standard normal deviates, of the random effect of the *jth* subject in the *ith* treatment *d* was generated as

$d = \sigma_{subj} Z$, where the vector *Z* was generated using NORMAL function and σ^2_{subj} set to equal to one. Then the 9 7x1 vectors of observation for model (2.2) were calculated as

$\mathbf{y}_{ij} = d + \Sigma^{\frac{1}{2}} \varepsilon_{ij}$; $i = 1,2,3; j = 1,2,3$ where:

$\Sigma^{\frac{1}{2}}$ is the Cholevsky decomposition of Σ, and

Σ is the covariance matrix of repeated measures effect "time".

Therefore, the vector $\mathbf{y}_{ij}$ s defined as follow in the context of the mixed model (2.2):

$\mathbf{y}_{ij} = \left[y_{ij1} \quad y_{ij2} \quad \ldots \quad y_{ija} \right]' \approx MVN(\boldsymbol{\mu}_i, \Sigma_{aa} = J\sigma^2_{subj} + \Sigma)$; where μ_i is the vector of means at *kth* time for the *ith* treatment, i.e. $\boldsymbol{\mu}_i = \left[\mu_{i1} \quad \mu_{i2} \quad \ldots \quad \mu_{ia} \right]$ and $\mu_{ik} = \alpha_i + \tau_k + (\alpha\tau)_{ik}$ where

$i = 1,2,\ldots,t; j = 1,2,\ldots,r; k = 1,2,\ldots,a$

RESULTS AND DISCUSSION

Results:

Results under the Null Hypothesis for Balanced Data:

Due to space limitations, we present only part of the total simulation results of the 24 scenarios. The complete results are available from the author upon request. Table 3-5 summarizes results of the average of the empirical Type I error rate across the 2 investigated settings of each covariance structure, when data are simulated under the null hypothesis, for the eight covariance structure for the tests of fixed effects with the two sample size. Table 3-5 indicates that the right analysis approach provided a very good control of the nominal error probabilities for all the effects all across the eight covariance structure. Although the nominal level was controlled well with the right analysis approach and the unstructured analysis approach, slightly departures were observed with the split-plot analysis approach for certain covariance structure particularly for sample size 5. As expected, control of the nominal error probability improved with increasing sample size.

Results under the Null Hypothesis for Unbalanced Data:

Table 6-8 indicates that the right analysis approach provided a very good control of the nominal error probabilities for all the effects all across the eight covariance structure. Although the nominal level was controlled well with the right analysis approach and the unstructured analysis approach, slightly departures were observed with the split-plot analysis approach for certain covariance structure particularly for sample size 5. As expected, control of the nominal error probability improved with increasing sample size.

Table 3: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Treatment Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.053500	0.056875	0.056375	0.052875	0.05300	0.05450	0.0565	0.05575
Unstructured model	0.056000	0.055500	0.056750	0.053375	0.05375	0.05075	0.0565	0.05775
Split-plot model	0.053500	0.055500	0.056750	0.053375	0.05375	0.04900	0.0565	0.05775
	Sample Size: 10							
Right model	0.050125	0.052625	0.05575	0.050375	0.05175	0.05175	0.051	0.0540
Unstructured model	0.050125	0.051875	0.05275	0.051375	0.05150	0.05050	0.051	0.0525
Split-plot model	0.050125	0.051875	0.05275	0.051375	0.05150	0.05050	0.051	0.0525

Table 4: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Time Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.052875	0.041250	0.044250	0.047125	0.05025	0.04350	0.05400	0.05150
Unstructured model	0.042500	0.045875	0.039750	0.044875	0.03925	0.04625	0.05400	0.03975
Split-plot model	0.052875	0.086625	0.076125	0.096000	0.05125	0.07225	0.08625	0.07700
	Sample Size: 10							
Right model	0.048250	0.049000	0.042750	0.049375	0.04975	0.04725	0.05000	0.04600
Unstructured model	0.046375	0.051625	0.049250	0.047625	0.04800	0.04875	0.05000	0.04900
Split-plot model	0.048250	0.084500	0.077125	0.092125	0.04925	0.07575	0.08325	0.07775

Table 5: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Interaction Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.051500	0.049000	0.040375	0.050000	0.04925	0.04925	0.052	0.05525
Unstructured model	0.045125	0.049375	0.044500	0.050125	0.04550	0.04975	0.052	0.04425
Split-plot model	0.051500	0.097875	0.082375	0.107625	0.05100	0.07950	0.104	0.08425
	Sample Size: 10							
Right model	0.049750	0.043125	0.042875	0.040750	0.04475	0.04575	0.0510	0.05050
Unstructured model	0.047000	0.051375	0.047375	0.046750	0.05275	0.04900	0.0510	0.04825
Split-plot model	0.049750	0.087625	0.087750	0.107875	0.04550	0.07575	0.0935	0.08650

Table 6: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Treatment Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.053750	0.058500	0.0567500	0.050875	0.05125	0.05400	0.05575	0.05450
Unstructured model	0.053375	0.057375	0.0547431	0.053875	0.05725	0.05275	0.05575	0.05925
Split-plot model	0.053750	0.053375	0.0550000	0.052000	0.05500	0.05275	0.05500	0.05775
	Sample Size: 10							
Right model	0.049500	0.05275	0.055875	0.049625	0.05300	0.05075	0.05025	0.05300
Unstructured model	0.050250	0.051650	0.0532500	0.051750	0.05150	0.04800	0.05025	0.05050
Split-plot model	0.049500	0.048125	0.0536250	0.048625	0.05275	0.04575	0.04950	0.05225

Table 7: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Time Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.054875	0.045000	0.04400	0.050750	0.05100	0.04725	0.062	0.05375
Unstructured model	0.045625	0.049375	0.04850	0.049500	0.04850	0.04275	0.062	0.04600
Split-plot model	0.054875	0.082500	0.07675	0.091875	0.05075	0.07200	0.082	0.07775
	Sample Size: 10							
Right model	0.051125	0.047750	0.043500	0.048750	0.05150	0.04900	0.061	0.04850
Unstructured model	0.053125	0.053625	0.052500	0.052000	0.05125	0.04950	0.061	0.05225
Split-plot model	0.051125	0.077875	0.075125	0.086000	0.05000	0.07300	0.076	0.07550

Table 8: Average of the Empirical Type I Errors across the two Investigated Covariance settings for Interaction Effect Under the Null Hypothesis (nominal Type I error=0.05).

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.051125	0.049000	0.040625	0.048625	0.04775	0.05175	0.06850	0.06200
Unstructured model	0.045625	0.054500	0.049750	0.053250	0.05025	0.04450	0.06850	0.04925
Split-plot model	0.051125	0.091750	0.082500	0.103875	0.05100	0.08100	0.09550	0.08475
	Sample Size: 10							
Right model	0.050375	0.042875	0.042125	0.044000	0.04775	0.05000	0.05775	0.05150
Unstructured model	0.050125	0.052750	0.048250	0.049375	0.05275	0.05475	0.05775	0.05000
Split-plot model	0.050375	0.080875	0.089125	0.101125	0.04800	0.06900	0.08125	0.0885

Results under the Alternative Hypothesis for Balanced Data:

Table 9-11 showed that overall the empirical powers were highest for the right analysis approach for all effects. Also, they showed that the empirical power was higher for the split-plot analysis approach comparing to the unstructured analysis approach for certain covariance structure particularly for sample size 5 due to their higher Type I error rates. As expected, the empirical power improved with increasing sample size.

Results under the Alternative Hypothesis for Unbalanced Data:

Table 12-14 showed that overall the empirical powers were highest for the right analysis approach for all effects. Also, they showed that the empirical power was higher for the split-plot analysis approach comparing to the unstructured analysis approach for certain covariance structure particularly for sample size 5 due to their higher Type I error rates. As expected, the empirical power improved with increasing sample size.

Table 15 showed the average percentage number of times across the investigated settings of covariance structure that the PROC MIXED procedure failing to converge when the PROC MIXED procedure used REML without any interfering for the balanced data.

The result in table 15 indicted that the percentages of number of times that the PROC MIXED Procedure failing to converge when the PROC MIXED Procedure used REML without any interfering were much higher for the unstructured covariance matrices which may be due to the large number of parameters for this structure comparing to the other structures.

Finally, Table 16 showed the average percentage number of times across the investigated settings of covariance structure that the PROC MIXED procedure failing to converge when the PROC MIXED procedure used REML without any interfering for the unbalanced data.

The result in table 16 indicted that the percentages of number of times that the PROC MIXED Procedure failing to converge when the PROC MIXED Procedure used REML without any interfering were much higher for the unstructured covariance matrices which may be due to the large number of parameters for this structure comparing to the other structures. Also, the results in table 15 and 16 may suggest that the convergence problem could be more serious in the case of the unbalanced data comparing to the balanced data.

Table 9: Average of the Empirical Power across the two Investigated Covariance settings for Treatment Effect Under the Alternative Hypothesis.

The Analysis Approach	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Sample Size: 5								
Right model	0.053500	0.056875	0.056375	0.052875	0.05300	0.05450	0.0565	0.05575
Unstructured model	0.053500	0.055500	0.05675	0.053375	0.05375	0.05275	0.0565	0.05775
Split-plot model	0.053500	0.055500	0.05675	0.053375	0.05375	0.05275	0.0565	0.05775
Sample Size: 10								
Right model	0.050125	0.052625	0.05575	0.049750	0.05175	0.05175	0.0510	0.05400
Unstructured model	0.050125	0.051875	0.05275	0.051375	0.05150	0.05050	0.0510	0.05250
Split-plot model	0.050125	0.051875	0.05275	0.051375	0.05150	0.05050	0.0510	0.05250

Table 10: Average of the Empirical Power across the two Investigated Covariance settings for Time Effect Under the Alternative Hypothesis.

The Analysis Aapproach	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Sample Size: 5								
Right model	0.402375	0.211500	0.220625	0.227500	0.13500	0.13600	0.93650	0.21600
Unstructured model	0.221375	0.146375	0.14100	0.149500	0.08300	0.09700	0.93650	0.13125
Split-plot model	0.402375	0.187500	0.22825	0.158375	0.13450	0.11200	0.14400	0.23250
Sample Size: 10								
Right model	0.701625	0.488250	0.477875	0.498125	0.20725	0.29650	0.90650	0.43750
Unstructured model	0.608500	0.428000	0.401875	0.430000	0.20050	0.26200	0.90650	0.36475
Split-plot model	0.701625	0.312500	0.412875	0.227375	0.24975	0.16425	0.20675	0.41450

Table 11: Average of the Empirical Power across the two Investigated Covariance settings for Interaction Effect Under the Alternative Hypothesis.

The Analysis Aapproach	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Sample Size: 5								
Right model	0.196875	0.105750	0.104000	0.108125	0.08050	0.08225	0.72725	0.11175
Unstructured model	0.113625	0.083875	0.079875	0.085500	0.05750	0.06450	0.72725	0.07425
Split-plot model	0.196875	0.143125	0.147125	0.139250	0.08125	0.10025	0.12950	0.14475
Sample Size: 10								
Right model	0.401625	0.233750	0.218125	0.233500	0.1040	0.14325	0.90075	0.20425
Unstructured model	0.320250	0.206125	0.194375	0.204125	0.1060	0.12750	0.90075	0.17325
Split-plot model	0.401625	0.194375	0.221500	0.160750	0.1210	0.11125	0.15100	0.22300

Table 12: Average of the Empirical Power across the two Investigated Covariance settings for Treatment Effect Under the Alternative Hypothesis.

The Analysis Approach	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR (1)	ARH (1)	VC	UN (1)	UN	TOEP
Sample Size: 5								
Right model	0.053750	0.058500	0.05675	0.050875	0.05125	0.05400	0.0562359	0.05450
Unstructured model	0.053375	0.057250	0.05475	0.053750	0.05725	0.05075	0.0562359	0.05925
Split-plot model	0.053750	0.053375	0.05500	0.052000	0.05500	0.04900	0.0550000	0.05775
Sample Size: 10								
Right model	0.04950	0.05275	0.055875	0.049625	0.05300	0.05075	0.0502500	0.05300
Unstructured model	0.05025	0.051625	0.053250	0.051750	0.05150	0.04800	0.0502500	0.05050
Split-plot model	0.04950	0.048125	0.053625	0.048625	0.05275	0.04575	0.0495000	0.05225

Table 13: Average of the Empirical Power across the two Investigated Covariance settings for Time Effect Under the Alternative Hypothesis.

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.38500	0.204625	0.214875	0.223250	0.13475	0.13550	0.57725	0.2050
Unstructured model	0.19325	0.125250	0.12600	0.128750	0.07975	0.08925	0.57725	0.1175
Split-plot model	0.38500	0.178125	0.22025	0.149875	0.13700	0.11000	0.13325	0.2250
	Sample Size: 10							
Right model	0.67750	0.457375	0.44500	0.463375	0.19250	0.27900	0.72850	0.4100
Unstructured model	0.48725	0.384875	0.36475	0.379625	0.19125	0.23425	0.72850	0.3370
Split-plot model	0.67750	0.291875	0.39075	0.214625	0.23400	0.14925	0.18750	0.3930

Table 14: Average of the Empirical Power across the two Investigated Covariance settings for Interaction Effect Under the Alternative Hypothesis.

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0.183375	0.10275	0.101875	0.107750	0.07875	0.08125	0.38975	0.12075
Unstructured model	0.098875	0.081500	0.076000	0.079875	0.06350	0.06600	0.38975	0.07200
Split-plot model	0.183375	0.178125	0.139625	0.129500	0.08150	0.09450	0.12475	0.14075
	Sample Size: 10							
Right model	0.368000	0.216625	0.201125	0.216125	0.10025	0.13225	0.70750	0.18525
Unstructured model	0.284625	0.190375	0.191250	0.180750	0.10425	0.12425	0.70750	0.15100
Split-plot model	0.368000	0.174375	0.212125	0.148375	0.11400	0.09825	0.13500	0.30225

Table 15: The Percentage of number of times that the PROC MIXED Procedure failing to Converge when the PROC MIXED Procedure used REML without any interfering across the two Investigated Covariance settings of each Covariance structure.

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0%	0%	0%	0%	0%	0%	5.3%	0%
Unstructured model	0%	0%	0%	0%	0%	0.58%	5.3%	0%
Split-plot model	0%	0%	0%	0%	0%	0%	0%	0%
	Sample Size: 10							
Right model	0%	0%	0%	0%	0%	0%	12.1%	0%
Unstructured model	0%	0%	0%	0%	0%	0%	12.1%	0%
Split-plot model	0%	0%	0%	0%	0%	0%	0%	0%

Table 16: The Percentage of number of times that the PROC MIXED Procedure failing to Converge when the PROC MIXED Procedure used REML without any interfering across the two Investigated Covariance settings of each Covariance structure.

The Analysis Approach	Sample Size: 5							
	The Eight Covariance Structure used to simulate the data							
	CS	CSH	AR(1)	ARH(1)	VC	UN(1)	UN	TOEP
Right model	0%	0%	0%	0%	0%	0%	42.23%	0%
Unstructured model	6.42%	5.83%	5.355%	5.265%	0.65%	0%	42.23%	4.48%
Split-plot model	0%	0%	0%	0%	0%	0%	0%	0%
	Sample Size: 10							
Right model	0%	0%	0%	0%	0%	0%	33.8%	0%
Unstructured model	0.075%	0.14%	0.075%	0.14%	0%	0%	33.8%	0.03%
Split-plot model	0%	0%	0%	0%	0%	0%	0%	0%

Conclusion:

In our simulation, we considered repeated measure design, looking at the impact of restrict our options of covariance structure to the two options of the split-plot analysis approach and the unstructured analysis approach in both cases of balanced and unbalanced data. Overall, the right analysis approach provided the best power and in the same time it has very good control of the nominal Type I error rate for both cases of balanced and unbalanced data. Thus, this approach can be recommended to be used in both cases of balanced and unbalanced data with the Kenward- Roger method. The main result of our article is that the right analysis approach is superior in term of power comparing to the unstructured analysis approach when both the approaches yield satisfactory Type I error control with advantage of no convergence problem exist for the right analysis approach in both cases of balanced and unbalanced data. The no convergence problem with the unstructured analysis approach getting worse in case of unbalanced data comparing to case of balanced data for both the sample sizes. In addition, the empirical power was higher for the split-plot analysis approach comparing to the unstructured analysis approach for certain covariance structure due to their higher Type I error rates. As expected, overall the empirical powers were higher for the balanced data comparing to the unbalanced data for all the approaches when the powers improved for both cases of balanced and unbalanced data with increasing sample size.

REFERENCES

Bellavance, F., S. Tardif and M. Stephens, 1996. Tests for the analysis of cross-over designs with correlated errors. Biometrics, 52(2): 607-612.

Fai, A.H.T. and P.L. Cornelius, 1996. Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. J. Statist. Compu. and Simu., 54: 363-378.

Gill, P.S., 1992. Balanced change-over designs for auto correlated observations. Aust. J. Statist. 34(3): 415-420.

Giesbrecht, F.G. and J.C. Burns, 1985. Two stage Analysis Based on a Mixed Model: Large Sample Asymptotic Theory and Small-Sample Simulation Results. Biometrics, 41: 477-486.

Guerin, L. and W.W. Stroup, 2000. A simulation study to evaluate PROC MIXED Analysis of repeated measures data. In Proceedings of the 2000 Conference on Applied Statistics in Agriculture pp: 170-203. Manhattan, KS.: Kansas State University.

Henderson, C.R., 1963. Selection Index and Expected Genetic Advance. Statistical Gentics and Plant Breeding, NAS-NRC Publ. No. 982: 141-163.

Henderson, C.R., 1990. Statistical Method in Animal Improvement: Historical Overview. In Advances in Statistical Methods for Genetic Improvement of Livesstock, New York: Springer.

Huynh, H. and L.S. Feldt, 1970. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, 65: 1582-1589.

Jones, B. and M. Kenward, 1989. Design and Analysis of Cross-Over Trials. London: Chapman and Hall.

Kacker, R.N. and Harville, D.A., 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J. Amer. Statist. Assoc., 79: 853-862.

Kenward, M.G. and J.H. Rogers, 1997. Small sample inference for fixed effect from restricted maximum likelihood. Biometrics, 53: 983-97.

McCulloch, C. and S. Searle, 2001. Generalized, Linear, and Mixed Models. Wiley, New York.

SAS Institute, 2003. *SAS/STAT,* Release 9.1. Cary NC: SAS Institute Inc.

Satterthwaite, F.E., 1941. Synthesis of Variance. Psychometrika, 6: 353-367.