

POWER ANALYSIS TO DETERMINE THE IMPORTANCE OF COVARIANCE
STRUCTURE CHOICE IN MIXED MODEL REPEATED MEASURES ANOVA

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By
Taylor J King

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

November 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Power Analysis to Determine the Importance of Covariance Structure
Choice in Mixed Model Repeated Measures Anova

By

Taylor J King

The Supervisory Committee certifies that this *disquisition* complies with North Dakota
State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel

Chair

Dr. Curt Doetkott

Dr. Seung Won Hyun

Dr. Simone Ludwig

Approved:

11/15/2017

Date

Dr. Rhonda Magel

Department Chair

ABSTRACT

Repeated measures experiments involve multiple subjects with measurements taken on each subject over time. We used SAS to conduct a simulation study to see how different methods of analysis perform under various simulation parameters (e.g. sample size, autocorrelation, repeated measures). Our goals were to: compare the multivariate analysis of variance method using PROC GLM to the mixed model method using PROC MIXED in terms of power, determine how choosing the incorrect covariance structure for mixed model analysis affects power, and identify sample sizes needed to produce adequate power of 90 percent under different scenarios. The findings support using the mixed model method over the multivariate method because power is generally higher when using the mixed model method. Simpler covariance structures may be preferred when testing the within-subjects effect to obtain high power. Additionally, these results can be used as a guide for determining the sample size needed for adequate power.

TABLE OF CONTENTS

ABSTRACT.....	iii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
LIST OF APPENDIX TABLES.....	viii
LIST OF APPENDIX FIGURES.....	ix
CHAPTER 1. INTRODUCTION.....	1
CHAPTER 2. LITERATURE REVIEW.....	5
2.1. Mixed Model for Repeated Measures.....	5
2.2. Repeated Measures Analysis in SAS.....	7
2.3. Repeated Measures Simulation Research.....	10
CHAPTER 3. SIMULATION METHODOLOGY.....	14
3.1. Data Simulation.....	14
3.2. Simulation Parameters and Finding Power.....	16
CHAPTER 4. RESULTS.....	19
4.1. Checking Type I Error Rates.....	19
4.2. Power for Treatment Effect.....	25
4.3. Power for Time Effect.....	30
4.4. Power for Interaction Effect.....	36
CHAPTER 5. MOTIVATING EXAMPLE.....	42
CHAPTER 6. CONCLUSION.....	49
REFERENCES.....	53
APPENDIX A. SIMULATION CODE.....	57

APPENDIX B. TYPE I ERROR RATES AND FIT STATISTICS.....65

APPENDIX C. CODE USED FOR EXAMPLE.....77

APPENDIX D. FURTHER POWER ANALYSIS AS CORRELATION INCREASES.....83

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Type I Error Rates for Treatment Effect.....	22
2. Type I Error Rates for Time Effect.....	23
3. Type I Error Rates for Interaction Effect.....	24
4. Power for Detecting an Effect Size of $(1/2)\sigma_s$ for Treatment Effect.....	29
5. Power for Detecting an Effect Size of σ_s for Treatment Effect.....	29
6. Power for Detecting an Effect Size of $2\sigma_s$ for Treatment Effect.....	30
7. Power for Detecting an Effect Size of $(1/2)\sigma_s$ for Time Effect.....	35
8. Power for Detecting an Effect Size of σ_s for Time Effect.....	35
9. Power for Detecting an Effect Size of $2\sigma_s$ for Time Effect.....	36
10. Power for Detecting an Effect Size of $(1/2)\sigma_s$ for Interaction Effect.....	40
11. Power for Detecting an Effect Size of σ_s for Interaction Effect.....	40
12. Power for Detecting an Effect Size of $2\sigma_s$ for Interaction Effect.....	41
13. Fit Statistics for Different Covariance Structures.....	45
14. P-values by Covariance Structure.....	48

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Repeated Measures Experiment when Number of Subjects equals 10.....	14
2. Example of PROC MIXED code when TYPE = AR(1).....	18
3. Example of PROC GLM code with 4 Repeated Measures	18
4. Treatment Effect – Trend of Power by Covariance Structure.....	25
5. Treatment Effect – Trend of Power by Number of Repeated Measures.....	26
6. Treatment Effect – Trend of Power by Correlation.....	26
7. Treatment Effect – Trend of Power by Effect Size.....	28
8. Time Effect – Trend of Power by Covariance Structure.....	31
9. Time Effect – Trend of Power by Number of Repeated Measures.....	32
10. Time Effect – Trend of Power by Correlation.....	33
11. Time Effect – Trend of Power by Effect Size.....	34
12. Interaction Effect – Trend of Power by Covariance Structure.....	37
13. Interaction Effect – Trend of Power by Number of Repeated Measures.....	38
14. Interaction Effect – Trend of Power by Correlation.....	38
15. Interaction Effect – Trend of Power by Effect Size.....	39
16. SAS Code to Obtain the Estimated Correlation Matrix using Unstructured.....	43
17. SAS Output for the Estimated Covariance Matrix using Unstructured.....	43
18. SAS Output for the Estimated Correlation Matrix using Unstructured.....	43
19. Plot of Covariance of Between-Subjects vs Lag in Weeks for UN.....	44
20. Plot of Covariance of Between-Subjects vs Lag in Weeks for AR(1).....	46
21. Plot of Covariance of Between-Subjects vs Lag in Weeks for CS.....	46
22. Plot of Covariance of Between-Subjects vs Lag in Weeks for TOEP.....	47

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
B1. Type I Error for Time Effect when Effect Size on Treatment is $(\frac{1}{2})\sigma_s$	65
B2. Type I Error for Interaction Effect when Effect Size on Treatment is $(\frac{1}{2})\sigma_s$	65
B3. Type I Error for Time Effect when Effect Size on Treatment is σ_s	66
B4. Type I Error for Interaction Effect when Effect Size on Treatment is σ_s	66
B5. Type I Error for Time Effect when Effect Size on Treatment is $2\sigma_s$	67
B6. Type I Error for Interaction Effect when Effect Size on Treatment is $2\sigma_s$	67
B7. Type I Error for Treatment Effect when Effect Size on Time is $(\frac{1}{2})\sigma_s$	68
B8. Type I Error for Interaction Effect when Effect Size on Time is $(\frac{1}{2})\sigma_s$	68
B9. Type I Error for Treatment Effect when Effect Size on Time is σ_s	69
B10. Type I Error for Interaction Effect when Effect Size on Time is σ_s	69
B11. Type I Error for Treatment Effect when Effect Size on Time is $2\sigma_s$	70
B12. Type I Error for Interaction Effect when Effect Size on Time is $2\sigma_s$	70
B13. Type I Error for Treatment Effect when Effect Size on Interaction is $(\frac{1}{2})\sigma_s$	71
B14. Type I Error for Time Effect when Effect Size on Interaction is $(\frac{1}{2})\sigma_s$	71
B15. Type I Error for Treatment Effect when Effect Size on Interaction is σ_s	72
B16. Type I Error for Time Effect when Effect Size on Interaction is σ_s	72
B17. Type I Error for Treatment Effect when Effect Size on Interaction is $2\sigma_s$	73
B18. Type I Error for Time Effect when Effect Size on Interaction is $2\sigma_s$	73
B19. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0.....	75
B20. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0.25.....	75
B21. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0.75.....	76

LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
D1. Power for Time 1 vs Time 2, 3, and 4 – $k = 4$	83
D2. Power for Time Effect – $k = 4$	83
D3. Power for Time 1 vs Time 2, 3, 4, 5, and 6 – $k = 6$	84
D4. Power for Time Effect – $k = 6$	84
D5. Code to Obtain P-values for Pairwise Comparisons.....	85

CHAPTER 1. INTRODUCTION

There are many different fields of study that use repeated measures in research, such as animal science, pharmaceuticals, sports medicine, and psychology. Repeated measures experiments involve multiple subjects where measurements are taken on each subject over time. Often, repeated measures experiments follow a completely randomized design where the subjects are randomly assigned to different treatments. In repeated measures studies such as these, the treatment effect, time effect, and the interaction effect between treatment and time may all be of interest. A good example of a repeated measures experiment with treatment and time effects may be a pharmaceutical company testing the effects of a drug on blood pressure. The subjects of this experiment would be randomly assigned to a placebo or treatment group, and then each subject would have their blood pressure taken over a series of time intervals, generally equally spaced.

Repeated measures experiments have historically been analyzed using three different statistical methods. These methods consist of the traditional univariate analysis of variance, multivariate analysis of variance, and mixed models. SAS, a statistical software, has many built in procedures that enable researchers to perform these three statistical methods on repeated measures data (SAS Institute Inc., 2017). SAS was released in 1976, and univariate analysis of variance was the only method available in SAS at the time to perform repeated measures analysis (Littell, 2011). This analysis can be performed using PROC GLM in SAS. One issue with using the traditional univariate analysis to analyze repeated measures data is that it does not consider the possible within-subjects factor when blocking by subject is not used. The measurements taken on the same subject over time tend to be correlated, therefore the usual assumption of independent errors does not hold. For this reason, the results given by this method are generally not reliable.

In 1984, SAS improved the ability to analyze repeated measures data by adding the REPEATED statement to PROC GLM (Littell, Henry, & Ammerman, 1998). This statement allows for the use of the multivariate analysis of variance method because the repeated variable can be specified. This method can only model the data to have an unstructured covariance structure, but how severely the model assumptions may be violated can be assessed.

The last of the three methods is the mixed model method, which can be performed through PROC MIXED in SAS. The MIXED procedure was added to SAS in 1992 (Littell, 2011). One advantage of the mixed model method over the univariate and multivariate methods, is that it is the only method that allows for analysis on repeated measures data that has missing data values from individual subjects. The univariate method simply ignores the missing observations, and the multivariate method simply ignores the subjects with missing observations. Additionally, the MIXED procedure allows for specification of the repeated variables, random effects variables, and the covariance structure. By choosing the covariance structure, researchers can take the within-subjects correlations into account. However, choosing the correct covariance structure may be difficult.

The univariate analysis of variance method seems to be the least favorable method of analysis on repeated measure data. However, the multivariate analysis of variance method and the mixed model method are both used often, with some researchers preferring the multivariate method, and other preferring the mixed model method. For this reason, only these two methods will be considered further in this paper.

Along with the debate of which method of analysis is best for repeated measures data, there is also an issue regarding the power of such analyses. Most repeated measures studies have a small sample size due to the experiments or research that is being conducted. Additionally,

subjects may be dropped or lost from the study due to various reasons, one being that the researcher can no longer get in contact with a subject after time has passed. A small sample size often results in low statistical power no matter what method is being used. Consequently, due to low power, some studies that state a significant effect in their research, may not be able to be replicated to get the same results. This low power and low reliability in studies is an issue, and more insight on the sample size required for adequate power, defined as 90 percent, for repeated measures data is needed.

In this study, these issues were explored more in depth through repeated measures data sets that were simulated and then tested. Simulations were used to create many samples of repeated measures data with a specific covariance structure and various simulation parameters such as sample sizes, number of repeated measures, effect sizes, and correlations. Then this data was analyzed using PROC GLM with the REPEATED statement for a multivariate analysis of variance method, as well as with the PROC MIXED to perform the mixed model method. These results were then compared to determine which method performs better. The data was also analyzed after specifying various incorrect covariance structures in PROC MIXED to determine the consequences on power. These comparisons were made on data simulated from multiple samples sizes to determine which sample size produces adequate power in detecting various effect sizes. This study represents an extension of a study done by Guerin and Stroup (2000) who used simulations to investigate Type I error and selection criteria of covariance structures for repeated measures. Simply stated, the goals of this study are to answer the following questions:

- How does the performance of the multivariate analysis of variance method compare to the performance of the mixed model method with respect to power and Type I error?
- How does choosing the incorrect covariance structure affect the power?

- What sample sizes are needed to produce adequate power of 90 percent for datasets with different simulation parameters?

In the next section, the background of repeated measures analysis will be described more in depth, as well as past research. The methods and the setup of this simulation study will then be discussed. Following the methodology will be a section on the results from the simulations, along with some discussion on these results. Then, to show motivation behind the importance of the results, an example using real data will be shown. The last section of the paper will consist of the conclusion to wrap up the overall thoughts and findings of this simulation study.

CHAPTER 2. LITERATURE REVIEW

2.1. Mixed Model for Repeated Measures

Repeated measures studies are common in many different areas such as animal science, pharmaceuticals, psychology, and other similar disciplines. These studies generally follow a completely randomized design where subjects are randomly assigned to a treatment group and the response is then measured over a series of times for each subject. In this case, the treatment effect is the between-subjects effect, while the time effect is the within-subjects effect. The completely randomized design described above follows a mixed model, shown below in model (1).

$$y_{ijk} = \mu + \tau_i + \alpha_j + \gamma_{ij} + d_{k(i)} + e_{ijk} \quad (1)$$

Where y_{ijk} is the response variable measured on the k^{th} subject, in the i^{th} treatment group, on the j^{th} time. μ is the overall mean, τ_i is main effect for treatment, α_j is the main effect for time, and γ_{ij} is the interaction effect for treatment and time. $d_{k(i)}$ is the random effect for the k^{th} subject with the i^{th} treatment, and $d_{k(i)} \sim N(0, \sigma_S^2)$. e_{ijk} is the random error for the k^{th} subject, in the i^{th} treatment group, on the j^{th} time. $e_{ijk} \sim N(0, \sigma_R^2)$.

Model (1) can also be written in matrix form. Littell, Milliken, Stroup, Wolfinger, and Schabenberger (2006b) give good discussion and background on this model. The mixed model in matrix form is shown below in model (2).

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2)$$

Where \mathbf{X} is the design matrix for the fixed effects, $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{Z}\mathbf{u}$ denotes the random subject effects where \mathbf{Z} is the design matrix for the random effects and $\mathbf{u} \sim N(0, \mathbf{G})$, and $\mathbf{e} \sim N(0, \mathbf{R})$. In this paper, the marginal distribution of this mixed model was used to simulate the data. There will be more discussion on this in Simulation Methodology section.

The data for the mixed model described above generally do not fit the assumption that the errors are independent. The measurements taken on subjects over time tend to be correlated in repeated measures data. Therefore, to analyze the data, the correct covariance structure must be found.

There are multiple sources available that discuss the different types of covariance structures (Guerin & Stroup, 2000; Kincaid, 2005; Kowalchuk, Keselman, Algina & Wolfinger, 2004; Wolfinger, 1996). Five common covariance structures are described below, and are shown assuming the number of repeated measures, from here on referenced as k , is equal to four for example purposes.

Variance Components (VC) is the simplest covariance structure because only the diagonal needs to be estimated. The diagonal elements in the structure are equal to $\sigma_k^2 * 1$, which correspond to the k parameters for this covariance structure. The off-diagonal elements are assumed to be equal to zero. This structure is essentially assuming independent errors.

$$VC = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Compound Symmetry (CS) should be used when measures on the same subject have the same correlation. CS has two parameters, one being the subject correlation, and the other being the subject variance component, σ_s^2 .

$$CS = \sigma_s^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

First-Order Autoregressive (AR(1)) can only be used when times are equally spaced. The correlation for each pair of observations that are measured n units apart is equal to ρ^n . AR(1) has two parameters.

$$\text{AR}(1) = \sigma_s^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Toeplitz (TOEP) is similar to AR(1) because the measurements that are taken at closer times have similar correlations (Kincaid, 2005). TOEP has k parameters, which is more than CS and AR(1) so it is considered more complex than CS and AR(1).

$$\text{TOEP} = \sigma_s^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Unstructured (UN) is the most complex covariance structure because it has $k(k+1)/2$ parameters. These parameters have no restrictions since each variance can be different and each covariance can be different (Kowalchuk et al., 2004).

$$\text{UN} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

2.2. Repeated Measures Analysis in SAS

There are three different approaches for analyzing repeated measures data – univariate analysis, multivariate analysis, and mixed model analysis. All three of these approaches can be done in SAS. PROC GLM is used to perform both the univariate and the multivariate analysis, while PROC MIXED is used to perform the mixed model analysis.

The univariate analysis approach was the first method offered in SAS. This approach is done through PROC GLM by putting the treatment variable in the MODEL statement and the time variable in the RANDOM statement (Littell, Milliken, Stroup, Wolfinger & Schabenberger, 2006a). This approach requires that the covariance structure for the data has sphericity. A covariance structure has sphericity if the diagonal variance components are all equal, and the off-diagonal covariance components are all equal. The univariate method requires this because it does not take into consideration that the errors within the subjects might be more strongly correlated over time. Compound Symmetry satisfies the sphericity requirement. Field (2010) researched the possible issues that can occur when analyzing repeated measures data with a univariate approach, specifically when sphericity does not hold. Sphericity can be tested in PROC GLM using Mauchly's Test. When the test is not significant then we can assume sphericity. When the assumption of sphericity does not hold, the Greenhouse Geisser correction (GG) or the Huynh and Feldt correction (HF) can be used in PROC GLM to correct the degrees of freedom (Field, 2010).

Another option when sphericity does not hold is to use the multivariate approach in PROC GLM. This was the second approach available for the analysis of repeated measures in SAS when the REPEATED statement was added to PROC GLM. For the multivariate analysis, you put the treatment variable in the MODEL statement and the time variable in the REPEATED statement (Wolfinger & Chang, 1995). The multivariate analysis in PROC GLM assumes that the repeated measures data follows an unstructured covariance structure (Everitt, 1995). As discussed earlier, UN is the most complex covariance structure as it has the most parameters, so if the data truly follows a simpler covariance structure then the multivariate analysis may not be the most reliable because it will have low power. Wolfinger and Chang (1995) point out that in

PROC GLM, the analysis on the fixed effects is the same whether the univariate method is being used or the multivariate method is being used. The univariate and multivariate analysis only differ for the within-subjects factor, time. The model is estimated using the method of moments in PROC GLM. This does not allow for any missing observations. When just one observation value within a subject is missing, then that entire subject is thrown out for PROC GLM to estimate the model.

Of the three methods considered, the most recent addition to SAS that allows for analysis of repeated measures is PROC MIXED, which can perform the mixed model method of analysis. When using PROC MIXED, the model is estimated using the REML (restricted maximum likelihood) method. This method allows for missing observations in the dataset, which is one benefit of using PROC MIXED as opposed to PROC GLM. Another advantage of using PROC MIXED to analyze repeated measures data is that it allows for specification of the covariance in the TYPE option. The trouble, however, is that picking the correct covariance structure for the data can be difficult and require background knowledge about the behavior of the various covariance structures.

Of all three types discussed here, the multivariate analysis and the mixed model analysis are the most common as they allow for better handling of the within-subjects effect. The mixed model approach in SAS is more flexible, but requires more knowledge to be used correctly. The multivariate model is not as flexible, but is well understood and easy to use. These two approaches have been compared by multiple researchers, with most preferring PROC MIXED, (Gueorguieva & Krystal, 2004; Wang & Goonewardene, 2004). Stroup (1999) discusses the pros and cons of PROC MIXED, with an outline of cautions that need to be taken when using the mixed model approach. Due to the ongoing debate about which method is better, one goal of this

simulation study was to determine the difference between each method with respect to power and Type I error control.

2.3. Repeated Measures Simulation Research

There have been multiple papers published to help researchers better understand repeated measures analysis using PROC MIXED. Some of these papers describe the best way to efficiently and correctly code PROC MIXED for repeated measures analysis. When using PROC MIXED, there are two statements that can be specified, the RANDOM statement and the REPEATED statement. The TYPE option can be used in either the RANDOM or REPEATED statement to specify the covariance structure. It has been recommended to use PROC MIXED with the REPEATED statement when analyzing mixed models (Littell et al., 2006a). When only using the REPEATED statement, the TYPE option must be used in that statement to specify the covariance structure. Additionally, Kiernan, Tao, and Gibbs (2012) suggest that the use of the SUBJECT= option in the REPEATED statement can improve code efficiency. Mixed model analysis can be very time intensive, so it is important to have the most efficient code possible. In addition to being time intensive, PROC MIXED can also have convergence issues. Convergence can be an issue if the estimates are approaching too closely to zero. This issue can be solved by adjusting the lower bound (Guerin & Stroup, 2000). These suggestions were all considered when writing the code for this simulation study.

Choosing the correct covariance structure is another topic that has been researched for the mixed model method of analysis for repeated measures. Littell et al. (2006a) states that estimating the covariance structure for the data is one of the first steps in the analysis. Graphical tools to visualize the covariance structure are suggested, as well as looking at the various information criteria to determine the covariance structure with the best fit. A good example of

estimating the covariance structure for repeated measures data can be seen by Littell, Pendergast, and Naterajan (2000). It has also been suggested by researchers to use UN as the covariance structure when performing mixed model analysis on repeated measures data, especially when unsure which structure fits best (Kincaid, 2005; Kowalchuk et al., 2004). Although using UN has been suggested, it has also been determined that UN is not a suitable covariance structure when there is a small number of subjects compared to the number of repeated measures (Moser, 2004). Additionally, UN has many parameters, so loss of power can be expected. For these reasons, one objective of this simulation study was to determine the effects on power when testing repeated measures data after assuming the incorrect covariance structure. Results for when assuming UN were of special interest.

Along with literature regarding how to use SAS PROC MIXED, and determining the correct covariance structure for the analysis, there are also various simulation studies that have been conducted. Most of these simulation studies are done to determine the control over Type I error with repeated measures data under various distributions with different means and variances, as well as with various sample sizes and correlations.

Oberfeld & Franke (2013) ran simulations to determine Type I error control for repeated measures data with small sample size and non-normality in the response variable. They determined that the tests to analyze repeated measures are not appropriate to use when the data is not normal. There has also been research on controlling Type I error and power analysis using the mixed model Satherthwaite F Test and the multivariate Welch-Test (Keselman, Algina, Kowalchuk, & Wolfinger, 1999) . From this research, Keselman et al. (1999) were not able to say that one test controlled Type I error and had stronger power better than the other, which still leaves the question about which approach to use.

Kowalchuk et al. (2004) simulated data with between-subjects effect and within-subjects effect with multiple different covariance structures, as well as with unequal group sizes. From their study, they concluded that the KR adjusted degrees of freedom control the otherwise inflated Type I errors of complex covariance structures, such as UN. Note that the KR adjusted degrees of freedom estimate the denominator degrees of freedom for the F test to allow for inference of the fixed effects when sample size is small (Kenward, M.G., Roger, J.H, 1997). Kowalchuk et al. (2004) also discussed the importance that the simulated Type I error estimates should stay within a certain interval about alpha. Two possible intervals were mentioned in their paper – the binomial standard error interval, and the liberal criterion of robustness. If using the binomial standard error to calculate this interval, then the Type I error values should stay within $2 * \left[\frac{\alpha(1-\alpha)}{N} \right]^{1/2}$, where N is the total number of samples. The liberal criterion of robustness states the Type 1 error estimates should stay within 0.5α to 1.5α in order to be considered controlled (Bradley, 1978). For the simulation study conducted in this paper, these intervals were considered when checking the Type I error estimates.

Guerin and Stroup (2000) also conducted a simulation study to determine the Type I error for repeated measured data. They too found that using the KR adjustment in PROC MIXED can help avoid the overly inflated Type I errors that result from models with complex covariance structures. Additionally, Guerin and Stroup (2000) analyzed four model selection criteria: Akaike Information Criteria (AIC), Schwarz Bayesian Information Criteria (BIC), Corrected Akaike Information Criteria (AICC), and Hanan & Quinn Information Criteria (HQIC) to determine which criteria led to choosing the correct covariance structure most often. Guerin and Stroup (2000) found that the AIC is the best criterion to use when choosing a model because it

chooses a slightly more complex model, but not too complex of a model that power could be greatly impacted.

The set-up and methods behind the simulation study conducted in this paper were motivated by the simulation study performed by Guerin and Stroup (2000). The simulation study in this paper is mainly concerned with assessing the power of repeated measures analysis. It is known that power of repeated measures analysis can be an issue since these experiments are often conducted with a small sample size, but there is a lack of literature on the power of repeated measures analysis. The simulation study in this paper tries to fill the gap of knowledge on this subject.

CHAPTER 3. SIMULATION METHODOLOGY

3.1. Data Simulation

The simulation study performed in this paper involved simulating data from the marginal distribution of the mixed model shown in model (2) of the previous chapter assuming a completely randomized design with equally spaced time intervals. Every simulation was run with 5000 samples, 2 treatment groups, an AR(1) covariance matrix, a residual variance component (σ_R^2) equal to 1, a subject variance component (σ_S^2) equal to 4, and a mean equal to 5. The changing simulation parameters consisted of a various number of sample sizes, a various number repeated measurements, various effect sizes, and different correlations for the AR(1) covariance matrix.

The combinations of these different parameters, along with how the power for each simulation was attained, will be discussed next in section 3.2. First, a detailed walk through of how the data was simulated will be discussed and shown with an example. One of the simulations run in this study used data that was generated with 5 subjects per treatment group, 4 repeated measurements, and correlation equal to 0.25. This specific simulation is used in the following paragraphs to show examples of the generated data, and can be visualized in Figure 1.

Subject	1	2	3	4	5	6	7	8	9	10
Time Trt	1	1	1	1	1	2	2	2	2	2
1										
2										
3										
4										

Figure 1. Repeated Measures Experiment when Number of Subjects equals 10

The marginal distribution of model (2) is such that $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$ (Littell et al., 2006b). There are four main parts that had to be specified to obtain \mathbf{Y} . The first part consisted of generating the variance. Littell et al. (2006b) describe that one way of obtaining \mathbf{V} is to get rid of \mathbf{Z} and \mathbf{G} and set \mathbf{R} to be a block matrix with each block corresponding to a subject. This approach was taken in this study. Therefore, our initial covariance matrix (\mathbf{R}_0) was as

$$\text{follows, } \mathbf{R}_0 = \begin{bmatrix} 1 & 0.25 & 0.0625 & 0.015625 \\ 0.25 & 1 & 0.25 & 0.0625 \\ 0.0625 & 0.25 & 1 & 0.25 \\ 0.015625 & 0.0625 & 0.25 & 1 \end{bmatrix}.$$

From here, this matrix was used as the block matrix for each subject, and then the block matrix was multiplied by the subject variance component, σ_s^2 , which was set to equal 4 for this simulation study. The final \mathbf{R} matrix, which represents the $\text{Var}(\mathbf{Y})$, is then equal to 40 by 40 matrix to correspond to each of the 40 observations for each sample. ((2)(5)(4) = 40 rows because our example simulation had 2 treatment groups, 5 subjects per group, and 4 repeated measurements per subject.)

This \mathbf{R} matrix was next used to obtain the subject random error. The subject random error was generated to follow a normal distribution with a mean of zero, and a variance of \mathbf{R} . For this simulation, the subject error matrix had 40 rows corresponding to each observation, and 5000 columns corresponding to each sample.

The third part that had to be specified is $\mathbf{X}\boldsymbol{\beta}$, where \mathbf{X} is the design matrix and $\boldsymbol{\beta}$ is the vector of parameters. The vector $\boldsymbol{\beta}$ was set such that $\boldsymbol{\beta}^T = (\mu, \tau_1, \tau_2, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (5, 0, 0, 0, 0, 0, 0)$ when there is no effect size being simulated. The design matrix is then simulated to have

7 columns that correspond to each parameter in β , and 40 rows which correspond to each

observation. Therefore, the design matrix is as follows. $X =$

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This $X\beta$ portion is then repeated 5000 times because that was the number of samples being simulated.

Lastly, the random error components were generated to follow a normal distribution with a mean of zero, and a variance of σ_R^2 , where σ_R^2 was set to equal 1. This vector was repeated 5000 times, and therefore was a matrix with 40 rows corresponding to each observation, and 5000 columns corresponding to each sample.

Y was then obtained by adding $X\beta$, the subject random error components, and the random error components. The described steps above were done using PROC IML in SAS. The specific code can be seen in the Appendix A.

3.2. Simulation Parameters and Finding Power

Like the study done by Guerin and Stroup (2000), there were multiple simulations run with certain parameters set to equal various values for this study. As stated in section 3.1, all simulations were run with 2 treatment groups, 5000 samples, and an AR(1) covariance structure. The values used for the simulation parameters are as listed below.

- Total number of subjects: $s = 10, 20, 30$
- Number of repeated measures: $k = 4, 6$
- Correlation used in AR(1) covariance structure: $\rho = 0, 0.25, 0.75$

- Effect size: effect size on treatment only with $(1/2)\sigma_S$, σ_S , $2\sigma_S$; effect size on time only with $(1/2)\sigma_S$, σ_S , $2\sigma_S$; effect size on interaction term only $(1/2)\sigma_S$, σ_S , $2\sigma_S$

From the diagonal of \mathbf{R} , we know that $\sigma_S^2 = 4$. Therefore $(1/2)\sigma_S = 1$, $\sigma_S = 2$, and $2\sigma_S = 4$. The effect sizes are specified in $\boldsymbol{\beta}$. For example, when specifying an effect size of σ_S on the treatment effect only and $k = 4$, then $\boldsymbol{\beta}^T = (5, 0, 2, 0, 0, 0, 0)$. When specifying an effect size of σ_S on the treatment effect only and $k = 6$, then $\boldsymbol{\beta}^T = (5, 0, 2, 0, 0, 0, 0, 0, 0)$. To generate realistic data when simulating an effect size on time, each time was increased by a constant amount. For example, for an effect size of σ_S^2 on the time effect when $k = 4$, $\boldsymbol{\beta}^T = (5, 0, 0, 0, 0.67, 1.33, 2)$. For an effect size of σ_S^2 on the time effect when $k = 6$, $\boldsymbol{\beta}^T = (5, 0, 0, 0, 0.4, 0.8, 1.2, 1.6, 2)$. The data were simulated using all combinations of s , k , and ρ for each effect size. This resulted in 18 simulated datasets per effect size, therefore a total of 162 simulated data sets were run in this study to analyze power.

Each of these data sets was then analyzed using PROC GLM to obtain the power for multivariate analysis method, as well as analyzed using PROC MIXED to obtain the power for analysis of the mixed model method. Example code for PROC GLM and PROC MIXED can be seen below in Figures 2 and 3, respectively. Note that in PROC MIXED the KR adjusted degrees of freedom was used. This is because the Type I error was not controlled for all covariance structures without it. Also, Kenward & Roger (1997) suggest to use their approach when samples are small, which is the case for this simulation study. The Type I error results for this simulations study are shown in the first section of the Chapter 4.

As shown, both PROC GLM and PROC MIXED are set up to run the tests by sample. From PROC GLM, the Wilk's Lambda p-value was obtained for each sample when testing for the time effect or treatment*time effect. The number of samples with p-values less than or equal

to 0.05 was divided by the total number of samples to determine power. The power for testing for the treatment effect size in PROC GLM was found by taking the number samples with Type III Sum of Squares ANOVA p-values less than or equal to 0.05 and dividing by the total number of samples.

```
proc mixed data= rm.uv_AR1_4_10;  
  by sample;  
  class trt time subj_id;  
  model y = trt | time / ddfm=kr;  
  repeated time / subject=subj_id type=AR(1);  
  title2 "Repeated Measures ANOVA using Mixed Model Approach  
  -- AR(1)";  
run;
```

Figure 2. Example of PROC MIXED code when TYPE = AR(1)

```
proc glm data=rm.mv_AR1_4_10;  
  by sample;  
  class trt;  
  model y1 - y4 = trt / nouni;  
  repeated time 4;  
  title2 'Repeated Measures ANOVA for Effect of Time on  
  Response';  
run;
```

Figure 3. Example of PROC GLM code with 4 Repeated Measures

In PROC MIXED, the significance for treatment, time, or treatment*time can all be found by looking at the p-values obtained from the F statistics in the Type III Sum of Squares ANOVA. Like with PROC GLM, the power was obtained by taking the number of samples with p-values less than or equal to 0.05, and dividing by the total number of samples.

From these power results, the trend in power was observed for each analysis which was used to try to determine which analysis is superior. Additionally, when the data sets were run in PROC MIXED, they were run with the TYPE option set to be equal to VC, CS, AR(1), TOEP, and UN. This was done to determine how choosing the incorrect covariance structure can impact the power. The results also shed light onto how large a sample size must be to have adequate power. Note that in this paper, a power of 90 percent or larger is considered adequate. The results for these simulations can be seen in the next chapter of this paper.

CHAPTER 4. RESULTS

For this study, simulations were run with different combinations of sample sizes, number of repeated measures, and different correlations. These simulated data sets were then analyzed to determine the power for both the multivariate and mixed model methods of analysis. The powers found will be displayed in this chapter. The first section of this chapter will show the Type I error rates to assess whether the rates are under control. Next, the power for various effect sizes on the treatment effect will be discussed. Section 3 of this chapter will show the power results for the various effect sizes on the time effect. Lastly, the power of the interaction effect will be discussed.

4.1. Checking Type I Error Rates

In order to conduct this power analysis, a Type I error rates analysis had to be conducted to show that all rates were controlled. The Type I error rates for all three methods and of analyzing repeated measures data are shown on Tables 1, 2, and 3 in percentage form. Note that the PROC GLM univariate analyses of Type I error are only included on these tables for the reader's information. There was no power analysis performed with the univariate method.

Table 1 displays the Type I error rates for the treatment effect. For most methods and specified covariance structures, the Type I error rates were kept within appropriate limits of alpha equal to 0.05. According to the binomial standard error interval, the Type I error rates for 5000 samples should stay within 0.0438 and 0.0562. According to the liberal criterion of robustness, Type I errors rates should stay within 0.025 and 0.075. The F test from PROC GLM is the result for the multivariate and univariate approach of the between-subjects effect. Recall from Chapter 2 of this paper that the multivariate method has the same between-subjects effect test as the univariate method, and assumes a compound symmetry covariance structure.

Therefore, the PROC GLM F test results for the between-subjects effect can be considered the Type I error rates for both the univariate and multivariate method. Note that the PROC MIXED analysis with CS specified as the covariance structure resulted in Type I error rates equal to the PROC GLM F test since the PROC GLM assumes CS as the covariance structure when testing the between-subjects effect.

Table 1 also shows the PROC GLM univariate analysis with no blocking and the PROC MIXED analysis with VC specified as the covariance structure. These two analyses did not result in controlled Type I error rates when correlation was greater than 0. The PROC GLM univariate method with no blocking does not consider the within-subjects effect, and therefore assumes all observations are independent. The VC covariance structure also assumes the observations within-subjects are not correlated by always assuming the correlation is equal to zero. Consequently, the error rates for these two analyses were controlled when the correlation equals 0. However, as ρ increased, the Type I error rates for these two situations become very inflated.

Table 2 displays the Type I error rates for testing the time effect. Here, the PROC GLM Wilk's Lambda test was used to obtain the multivariate method test, and did have controlled Type I error rates. The univariate F Test, GG, and HF are the univariate results obtained through the same PROC GLM code. Recall from Chapter 2 that the GG and HF tests adjust the degrees of freedom to better improve the accuracy of the original test for the univariate method. Table 2 shows that the PROC GLM univariate F Test results and the PROC GLM HF results had controlled Type I error rates. The PROC GLM GG analysis results are considered controlled as well, but note that the Type I errors rates were more conservative compared to the other tests. Once again, the PROC GLM univariate method with no blocking, and the PROC MIXED method with VC specified as the covariance both resulted in conservative error rates. When

correlation was set equal to 0, the error rates were controlled. As the correlation increased, the Type I error rates dramatically decreased.

The Type I error rates for the interaction effect are shown on Table 3. This table shows that once again the PROC GLM univariate analysis with no blocking, and the PROC MIXED test with VC specified as the covariance both resulted in uncontrolled error rates. The trends for these Type I errors were like the trends for the period effect in that as the correlation increased, the Type I error rates for these tests decreased.

Overall, the PROC GLM multivariate method analyses displayed controlled Type I error rates. The PROC MIXED analyses showed controlled Type I error for all covariance structures specified excluding the VC covariance structure. Therefore, the VC covariance structure was not analyzed further in the power analysis. The univariate analysis methods were not of interest in this study, but the Type I error results were included in this section to provide some helpful insight. The inflation in the Type I error rates for the PROC GLM univariate analyses with no blocking results show why caution should be used when assuming all observations of a repeated measures experiment are independent. The next sections will display the power analysis for the PROC GLM multivariate method, and the PROC MIXED method with CS, TOEP, UN, or AR(1) as the covariance structure.

Table 1. Type I Error Rates for Treatment Effect

Number of Samples = 5,000							
Binomial Proportions CI: (0.0438, 0.0562)							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM F test	10	4.86	4.82	4.94	5.16	4.84	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC GLM UV No Blocking	10	4.58	4.66	8.96	9.80	21.42	26.72
	20	5.04	4.76	9.46	10.08	21.44	26.94
	30	5.00	5.02	8.76	9.42	22.40	27.50
PROC MIXED VC	10	4.58	4.66	8.96	9.80	21.42	26.72
	20	5.04	4.76	9.46	10.08	21.44	26.94
	30	5.00	5.02	8.76	9.42	22.40	27.50
PROC MIXED CS	10	4.86	4.82	4.94	5.16	4.88	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC MIXED TOEP	10	4.76	4.92	4.88	5.42	4.84	4.76
	20	4.82	5.06	5.34	5.10	4.88	5.30
	30	5.08	4.78	5.02	4.58	5.44	4.82
PROC MIXED UN	10	4.86	4.82	4.94	5.16	4.88	4.38
	20	4.76	5.06	5.38	5.24	4.78	5.16
	30	5.10	4.74	5.04	4.50	5.52	4.70
PROC MIXED AR(1)	10	4.74	4.62	5.04	5.62	6.20	6.78
	20	4.80	5.16	5.52	5.18	5.74	6.88
	30	5.30	4.86	5.10	5.02	6.38	6.44

Table 2. Type I Error Rates for Time Effect

Number of Samples = 5,000							
Binomial Proportions CI: (0.0438, 0.0562)							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.78	5.12	5.04	5.76	4.82	4.86
	20	4.96	4.92	4.78	4.56	4.80	4.42
	30	4.34	4.88	4.76	5.06	5.00	4.58
PROC GLM F test	10	5.16	5.50	5.28	5.36	5.52	6.68
	20	4.60	4.32	4.86	5.36	5.20	6.36
	30	4.42	4.84	4.92	5.26	5.76	6.40
PROC GLM GG	10	3.36	2.56	3.60	2.44	3.30	3.04
	20	3.90	3.08	3.76	3.62	3.84	3.88
	30	3.86	3.82	4.24	3.90	4.66	4.34
PROC GLM HF	10	4.72	4.94	4.70	4.76	4.52	5.62
	20	4.44	4.16	4.58	4.84	4.72	5.44
	30	4.48	4.74	4.72	4.82	5.18	4.98
PROC GLM UV No Blocking	10	5.16	5.38	3.38	3.88	0.66	1.06
	20	4.64	4.44	3.14	3.66	0.36	0.68
	30	4.48	4.86	3.36	3.40	0.38	0.60
PROC MIXED VC	10	5.16	5.38	3.38	3.88	0.66	1.06
	20	4.64	4.44	3.14	3.66	0.36	0.68
	30	4.48	4.86	3.36	3.40	0.38	0.60
PROC MIXED CS	10	5.16	5.50	5.28	5.36	5.52	6.68
	20	4.60	4.32	4.86	5.36	5.20	6.36
	30	4.42	4.84	4.92	5.26	5.76	6.40
PROC MIXED TOEP	10	4.96	5.96	5.70	6.24	5.10	5.72
	20	5.16	4.58	4.80	4.58	4.86	4.86
	30	4.32	4.94	4.88	4.92	5.40	4.56
PROC MIXED UN	10	4.78	5.12	5.04	5.76	4.82	4.86
	20	4.96	4.92	4.78	4.56	4.80	4.42
	30	4.34	4.88	4.76	5.06	5.00	4.58
PROC MIXED AR(1)	10	5.04	5.44	4.98	4.88	3.62	3.70
	20	4.76	4.46	4.76	4.60	4.18	3.88
	30	4.28	4.76	4.64	4.84	4.62	3.94

Table 3. Type I Error Rates for Interaction Effect

Number of Samples = 5,000							
Binomial Proportions CI: (0.0438, 0.0562)							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.78	5.48	4.64	4.64	4.60	5.48
	20	4.70	4.86	5.28	5.48	5.14	4.94
	30	4.76	5.10	5.34	4.80	4.76	5.24
PROC GLM F test	10	4.50	5.44	5.44	5.94	5.44	6.50
	20	4.64	4.78	5.32	5.70	5.12	6.40
	30	4.92	5.20	5.46	5.34	5.34	6.44
PROC GLM GG	10	2.90	2.36	3.22	2.76	3.70	2.58
	20	3.80	3.18	4.32	3.86	3.98	3.98
	30	4.40	4.06	4.72	4.04	4.30	4.50
PROC GLM HF	10	4.18	4.82	4.78	5.24	5.06	5.24
	20	4.50	4.62	5.14	5.16	4.68	5.34
	30	4.74	4.96	5.20	5.02	4.74	5.30
PROC GLM UV No Blocking	10	4.80	5.52	3.44	4.54	0.76	0.92
	20	4.84	4.78	3.54	4.10	0.60	0.86
	30	4.88	5.10	3.50	3.62	0.30	0.82
PROC MIXED VC	10	4.80	5.52	3.44	4.54	0.76	0.92
	20	4.84	4.78	3.54	4.10	0.60	0.86
	30	4.88	5.10	3.50	3.62	0.30	0.82
PROC MIXED CS	10	4.50	5.44	5.44	5.94	5.44	6.50
	20	4.64	4.78	5.32	5.70	5.12	6.40
	30	4.92	5.20	5.46	5.34	5.34	6.44
PROC MIXED TOEP	10	4.62	5.56	5.04	6.40	5.24	5.44
	20	4.72	5.00	5.38	5.56	4.96	4.80
	30	4.58	5.10	5.24	5.22	5.12	5.28
PROC MIXED UN	10	4.78	5.48	4.64	4.64	4.60	5.48
	20	4.70	4.86	5.28	5.48	5.14	4.94
	30	4.76	5.10	5.34	4.80	4.76	5.24
PROC MIXED AR(1)	10	4.36	5.10	4.74	5.24	3.68	3.50
	20	4.64	4.70	5.02	5.48	4.24	3.92
	30	4.68	5.18	5.24	5.06	4.60	4.42

4.2. Power for Treatment Effect

The Tables 4, 5, and 6 display the findings for the power of detecting various effect sizes on the treatment effect. The power results from all three tables display multiple trends. One trend being that as sample size increased, the power also increased for all methods and covariance structures used. To visualize this trend, see Figure 4, which plots the power as sample size increases for each covariance structure when effect size = σ_S , $k = 4$ and $\rho = 0.75$. Additionally, Figure 4 shows that the behavior among all analyses seem to be the same for all covariance structures. The PROC MIXED AR(1) analysis does show slightly higher power, although not drastically.

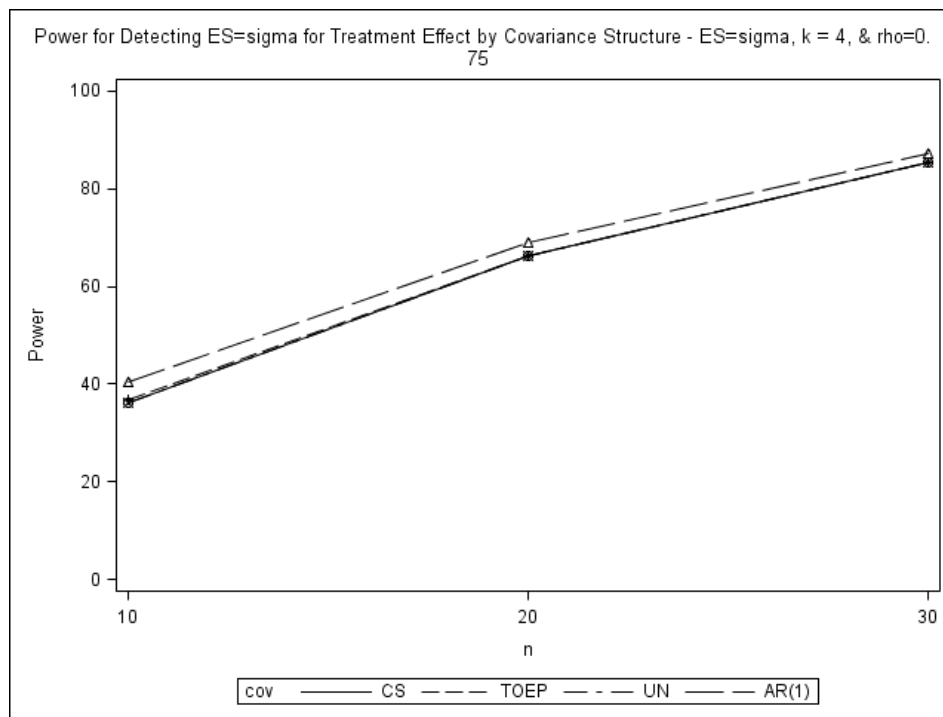


Figure 4. Treatment Effect – Trend of Power by Covariance Structure

Additionally, when looking across the columns of the tables, it can be noticed that as the number of repeated measures increases, the power also increases. To visualize the trend of power as repeated measures increase, see Figure 5. The powers on Figure 5 are from the PROC MIXED analysis when AR(1) was specified as the covariance structure, effect size = σ_S , and $\rho = 0.75$.

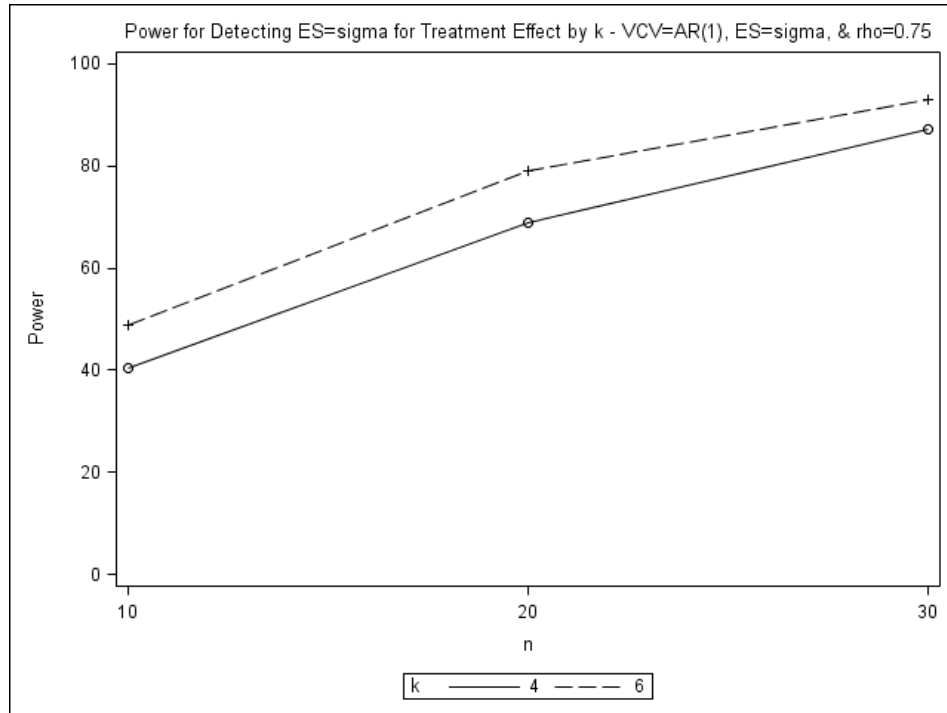


Figure 5. Treatment Effect – Trend of Power by Number of Repeated Measures

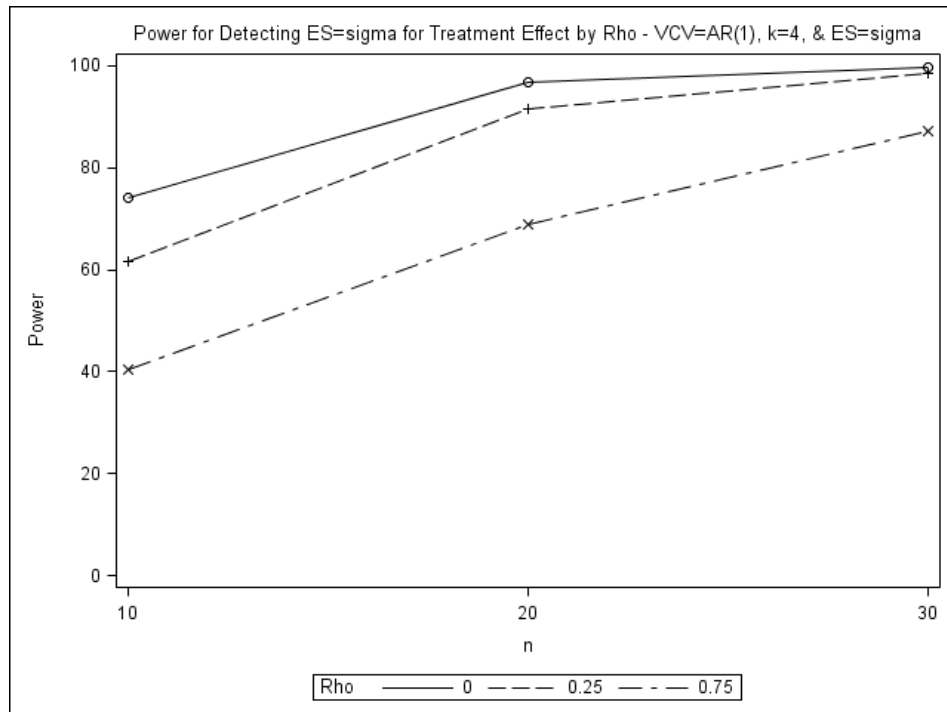


Figure 6. Treatment Effect – Trend of Power by Correlation

The power trend for correlation was different than the power trend for sample size and number of repeated measures. As the correlation increased, the power for detecting any effect

size on the treatment effect decreased. The variation of the treatment effect coefficient, found by determining $\text{Var}(\boldsymbol{\beta})$ using maximum likelihood estimation, increased as correlation increased which could explain this trend. To see this decreasing trend in power as correlation increases, view Figure 6. The power analysis on Figure 6 is from the PROC MIXED AR(1) analysis when $k = 4$ and effect size = σ_5 .

When comparing the power between the PROC GLM Wilk's Lambda analysis for the multivariate method and the PROC MIXED analysis for the mixed model method, it can be noticed that the power does not seem to differ by a large amount. The PROC MIXED results with AR(1) specified as the covariance structure consistently had the largest power, as mentioned earlier when referencing Figure 4. The data was simulated with an AR(1) covariance structure, so these results were expected.

The first effect size that the treatment effect was simulated with was $(1/2)\sigma_5$. These power results can be seen on Table 4. No combination of simulation parameters on this Table display a power of 90 percent or greater. Recall from Chapter 3 that adequate power was specified as 90 percent or higher. An effect size equal to σ_5 was simulated next, and the power results for this effect size can be seen on Table 5. These power results were much higher than for the effect size of $(1/2)\sigma_5$. A power of 90 percent or higher was obtained for a sample size of at least 20 with 4 repeated measures when the correlation was equal to 0. However, when the correlation was equal to 0.75, the sample size had to be at least 30, and the number of repeated measures had to be at least 6 to obtain a power of at least 90 percent.

When the data was simulated so that the treatment effect had an effect size of $2\sigma_5$, the power increased again. The power results for an effect size of $2\sigma_5$ on the treatment effect are in Table 6. When correlation was equal to 0, a power of 90 percent could be obtained with only a

sample size of 10, and 4 repeated measures. When the correlation was 0.75, a power of 90 percent could be obtained with a sample size of at least 20, and 4 repeated measures. The trends of power for each effect size can be visualized in Figure 7, which displays the power analysis results for PROC MIXED AR(1), when $k = 4$ and $\rho = 0.75$. Note that the Type I error rates for the time effect and the interaction effect remained controlled while simulating the data with the various treatment effect sizes. Tables B1, B2, B3, B4, B5, and B6 in Appendix B show these Type I error rates.

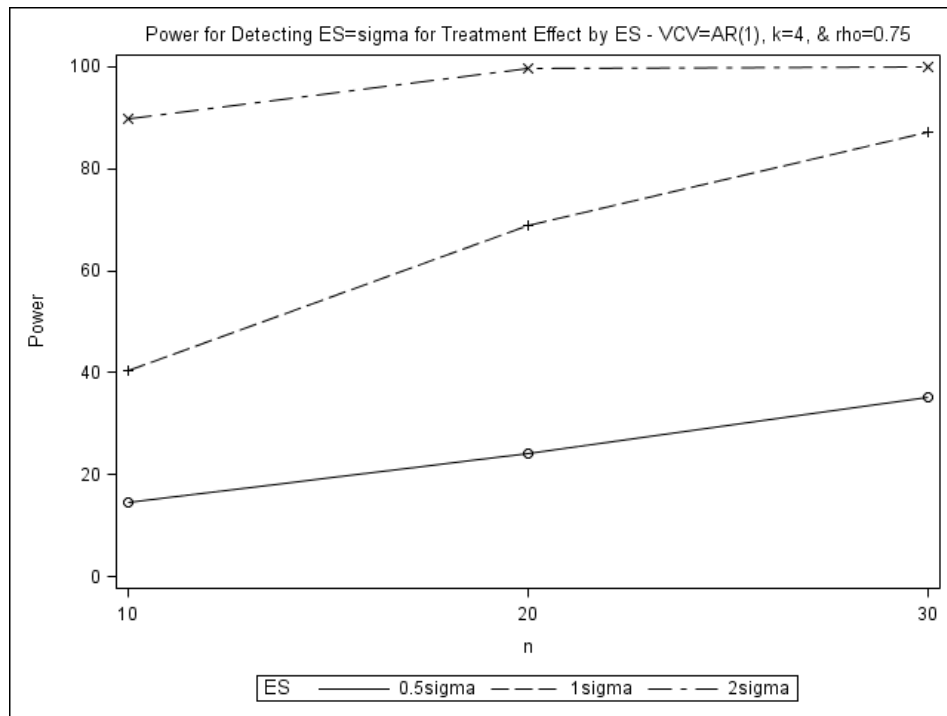


Figure 7. Treatment Effect – Trend of Power by Effect Size

Table 4. Power for Detecting an Effect Size of $(\frac{1}{2})\sigma_S$ for Treatment Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	22.62	34.10	19.04	24.60	11.74	12.82
	20	46.30	64.66	37.56	49.86	21.52	26.40
	30	65.32	82.46	51.98	68.22	32.06	36.92
PROC MIXED CS	10	22.62	34.10	19.04	24.60	11.74	12.82
	20	46.30	64.66	37.56	49.86	21.52	26.40
	30	65.32	82.46	51.98	68.22	32.06	36.92
PROC MIXED TOEP	10	22.94	34.18	19.24	24.94	12.32	13.18
	20	46.16	64.66	37.34	49.90	21.56	26.46
	30	65.42	82.46	52.20	68.00	32.26	36.92
PROC MIXED UN	10	22.62	34.10	19.04	24.60	11.74	12.82
	20	46.30	64.66	37.56	49.86	21.52	26.40
	30	65.32	82.46	51.98	68.22	32.06	36.92
PROC MIXED AR(1)	10	24.88	38.88	20.42	28.70	14.46	17.60
	20	47.86	66.70	39.30	52.36	24.14	30.84
	30	66.46	83.62	53.08	70.06	35.12	42.28

Table 5. Power for Detecting an Effect Size of σ_S for Treatment Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED CS	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED TOEP	10	69.68	85.88	57.98	71.62	36.76	41.52
	20	95.98	99.60	90.30	97.52	66.30	73.94
	30	99.74	100	98.26	99.76	85.48	90.62
PROC MIXED UN	10	70.32	86.36	58.04	72.16	36.14	41.10
	20	96.08	99.58	90.34	97.64	66.20	73.92
	30	99.74	100	98.34	99.78	85.40	90.34
PROC MIXED AR(1)	10	74.04	91.18	61.66	77.92	40.46	48.92
	20	96.90	99.72	91.48	98.06	69.02	78.96
	30	99.78	100	98.54	99.84	87.08	92.96

Table 6. Power for Detecting an Effect Size of $2\sigma_S$ for Treatment Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	99.90	100	98.92	99.80	87.10	92.52
	20	100	100	100	100	99.66	99.92
	30	100	100	100	100	100	100
PROC MIXED CS	10	99.90	100	98.92	99.8	87.10	92.52
	20	100	100	100	100	99.66	99.92
	30	100	100	100	100	100	100
PROC MIXED TOEP	10	99.80	99.96	98.78	99.8	87.10	92.24
	20	100	100	100	100	99.64	99.94
	30	100	100	100	100	100	100
PROC MIXED UN	10	99.90	100	98.92	99.8	87.10	92.52
	20	100	100	100	100	99.66	99.92
	30	100	100	100	100	100	100
PROC MIXED AR(1)	10	99.98	100	99.12	99.92	89.92	94.78
	20	100	100	100	100	99.78	99.96
	30	100	100	100	100	100	100

4.3. Power for Time Effect

The power results for detecting the various effect sizes on the time effect can be seen on Tables 7, 8, and 9. There were a few main trends that were realized from these power results. One trend, like the trend for the power in detecting the treatment effect size, is that as sample size increased, so did the power. A power plot to display the increase in power as sample size increased can be seen in Figure 8. The power analyses shown on Figure 8 display a line for each of the covariance structures specified in PROC MIXED when $k = 4$, $\rho = 0.75$, and effect size = σ_S .

The trend in power, however, did not increase as the number of repeated measures increased. Instead, the power was higher for 4 repeated measures than it was for 6 repeated measures. One possible explanation for this may be that the effect size is spread across more time periods when $k = 6$, which means the effect is incrementing by smaller amounts. This may

impact how well the effect size is detected for the time effect. This trend can be visualized in Figure 9, which uses the power analyses found from PROC MIXED AR(1), when $\rho = 0.75$ and effect size = σ_{ζ} .

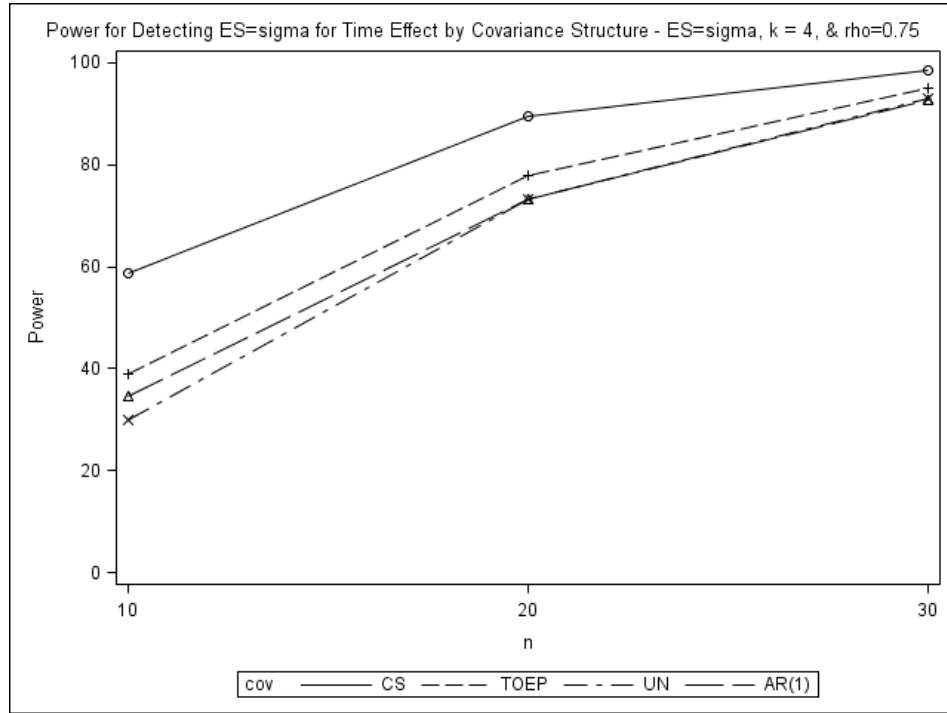


Figure 8. Time Effect – Trend of Power by Covariance Structure

There was not a consistent trend found as correlation increased that fit all covariance structures and methods that were tested. The PROC MIXED results with CS as the covariance structure, was the only test with power results that consistently increased as the correlation increased. This could possibly be due to the simplicity of the model, and the small sample sizes tested in this study. Most methods and covariance structures seemed to have a trend where the power would be higher for a correlation of 0 and 0.75, and a lower power when correlation was equal to 0.25. To visualize the trend of power for each correlation see Figure 10, which shows the power analyses of PROC MIXED AR(1) when $k = 4$ and effect size = σ_{ζ} . One possible explanation for the larger power when correlation is equal to 0, is that the covariance structures are not necessarily needed to estimate the model since a correlation of 0 implies independence.

To better understand the effect of correlation on the power of repeated measures data, pairwise comparisons of Time 1 vs Time 2, Time 1 vs Time 3, etc... were plotted. See Appendix D to reference these power trends further discovered with respect to correlation.

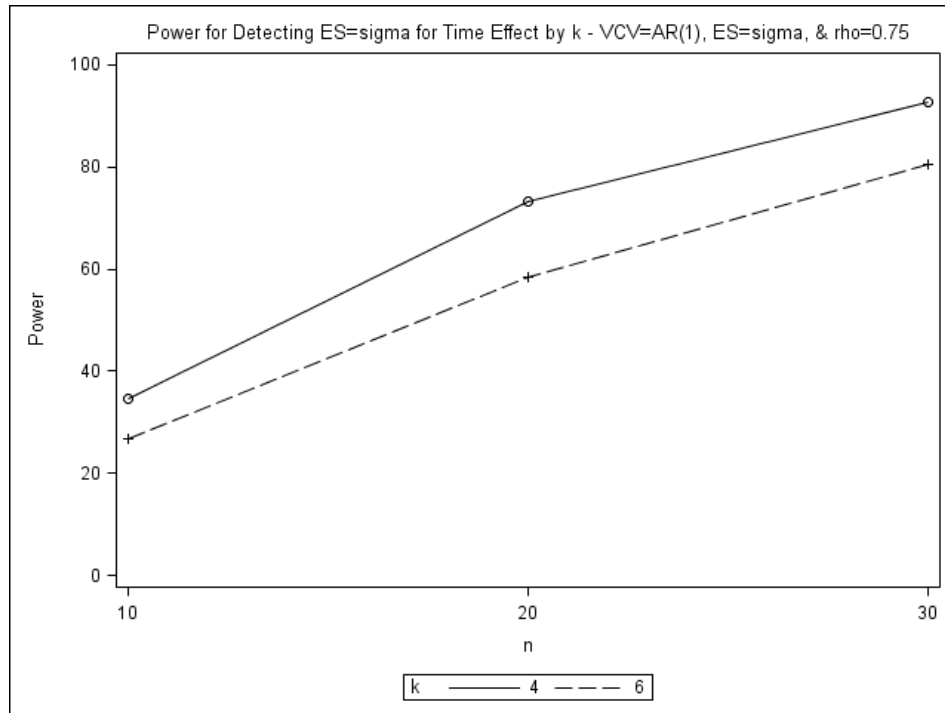


Figure 9. Time Effect – Trend of Power by Number of Repeated Measures

The power for the PROC MIXED results when UN is the specified covariance structure produced the lowest power trends compared to the other covariance structures. Notice these results are the same as the multivariate Wilk’s Lambda power, which is due to the KR degrees of freedom applied in PROC MIXED procedure. Recall that the PROC GLM multivariate method assumes the data has an unstructured covariance structure which also explains the similar results. The unstructured covariance structure is the most complex covariance structure used in this study, and therefore has a large number of parameters that need to be estimated, which can explain why it resulted in the lowest power. The PROC MIXED power results when CS is the specified covariance structure, produced the highest power compared to the other covariance

structures. Compound symmetry is one of the simpler structures that was tested, which may explain why it resulted in the highest power. This trend can be seen on Figure 8.

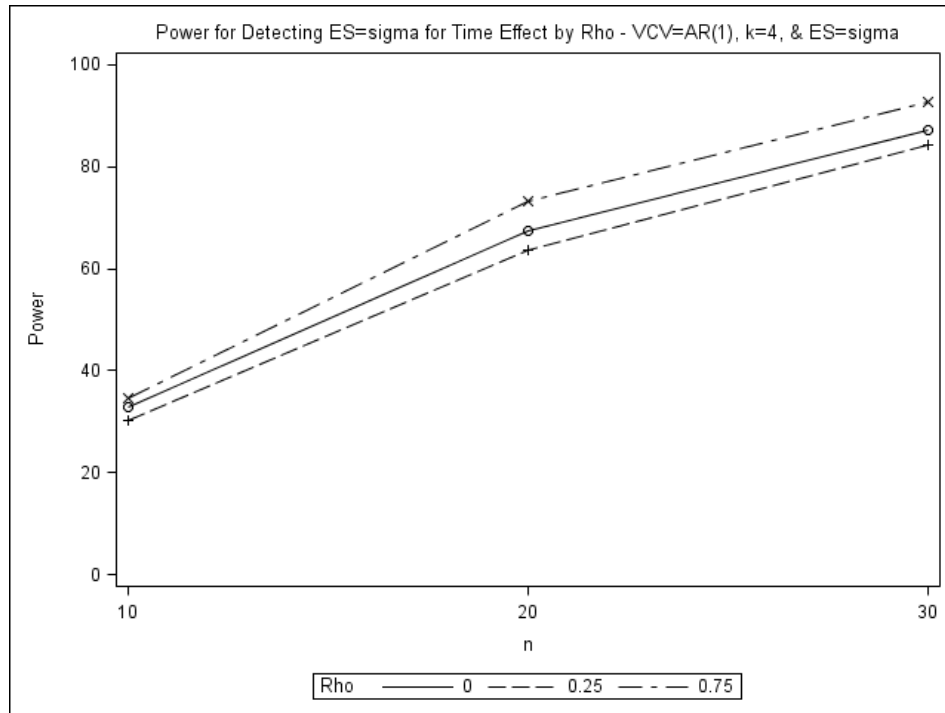


Figure 10. Time Effect – Trend of Power by Correlation

The power of detecting an effect size of $(\frac{1}{2})\sigma_S$ for the time effect is shown on Table 7. The results on Table 7 show that no combination of simulation parameters tested in this study produced a power of 90 percent for an effect size of $(\frac{1}{2})\sigma_S$ on the time effect. Therefore, a sample size of greater than 30 would be needed. This was not the case for the treatment effect, which means detecting a small effect size for the time effect may require a higher sample size than required for detecting a small effect size for the treatment effect.

The power results for an effect size of σ_S for the time effect is shown on Table 8. A power of at least 90 percent was consistently obtained for all methods and covariance structures when the sample size was at least 30, and the repeated measures was equal to 4. Therefore, detecting an effect size of σ_S or smaller for the time effect may require a sample size of at least 30 and $k = 4$. Table 9 displays the power results for an effect size of $2\sigma_S$. The power increased a

generous amount when compared to the power results for an effect size of σ_5 . A power of at least 90 percent was obtained for all combinations of simulation parameters and methods when the sample size was equal to at least 20. The trends for the power of the various effect sizes on the time effect can be seen on Figure 11. The power trends on Figure 11 display the results for each effect size applied to the time effect from the PROC MIXED AR(1) analyses when $k = 4$ and $\rho = 0.75$.

Overall, the larger the effect size, the higher the power. Like stated before, PROC MIXED CS produced the largest power. TOEP and AR(1) seem to behave similarly except when correlation is equal to 0.75 and AR(1) takes a small dip in power. Generally, PROC MIXED UN and PROC GLM Wilk's Lambda had the lowest power. As these power results were being tested on the time effect, the treatment and period effect Type I errors were controlled around alpha of 0.05. See Tables B7, B8, B9, B10, B11, and B12 of Appendix B for these Type I errors.

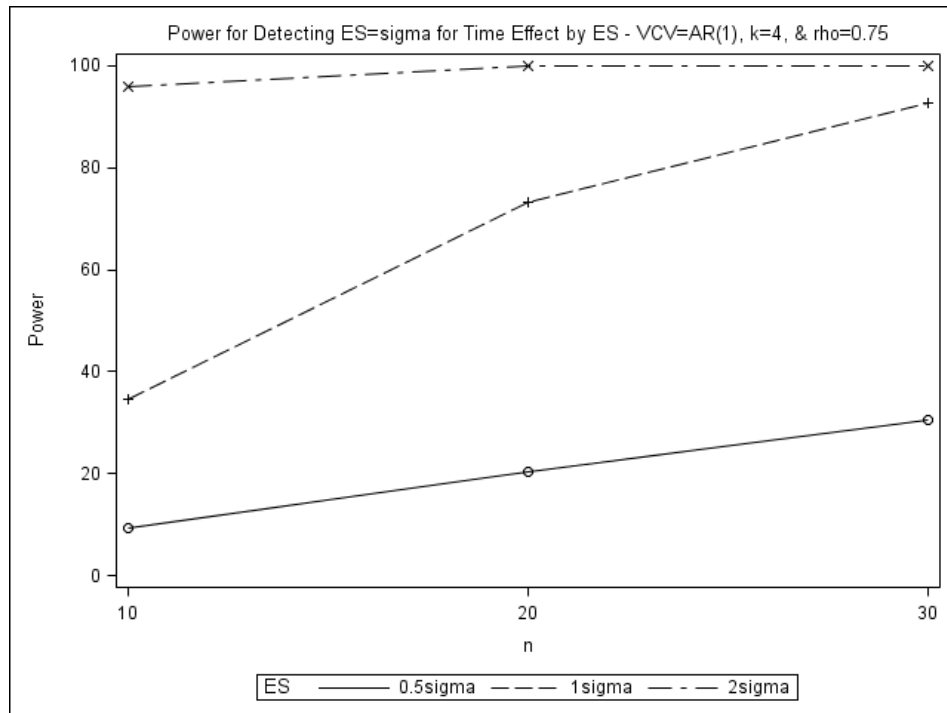


Figure 11. Time Effect – Trend of Power by Effect Size

Table 7. Power for Detecting an Effect Size of $(\frac{1}{2})\sigma_S$ for Time Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	8.92	8.00	8.92	6.80	10.44	6.72
	20	17.32	17.36	17.30	13.78	23.02	15.14
	30	27.20	25.98	25.26	22.44	35.08	24.48
PROC MIXED CS	10	11.60	11.52	12.66	12.68	19.32	19.32
	20	19.24	22.68	22.48	22.36	36.76	34.00
	30	29.20	30.98	32.68	34.10	50.06	48.82
PROC MIXED TOEP	10	10.56	11.52	10.52	10.18	12.58	10.34
	20	18.54	21.12	18.36	16.80	25.24	17.92
	30	29.08	29.70	27.30	25.78	36.18	28.00
PROC MIXED UN	10	8.92	8.00	8.92	6.80	10.44	6.72
	20	17.32	17.36	17.30	13.78	23.02	15.14
	30	27.20	25.98	25.26	22.44	35.08	24.48
PROC MIXED AR(1)	10	10.58	11.08	10.08	10.02	9.28	7.70
	20	19.12	22.52	18.50	16.82	20.28	14.94
	30	29.02	31.08	28.12	26.84	30.50	22.56

Table 8. Power for Detecting an Effect Size of σ_S for Time Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	22.94	16.88	20.72	13.72	30.00	14.14
	20	58.88	54.20	56.42	46.78	73.34	50.20
	30	82.86	82.62	79.46	73.30	93.04	79.16
PROC MIXED CS	10	33.64	37.18	37.16	38.46	58.74	55.12
	20	67.18	70.06	70.10	71.56	89.62	85.10
	30	87.34	89.64	88.16	88.18	98.42	95.90
PROC MIXED TOEP	10	30.12	31.86	28.38	26.30	39.12	29.60
	20	64.46	65.74	61.82	58.04	77.78	61.36
	30	86.10	87.62	82.50	79.94	94.90	84.06
PROC MIXED UN	10	22.94	16.88	20.72	13.70	30.00	14.14
	20	58.88	54.20	56.42	46.78	73.34	50.20
	30	82.86	82.62	79.46	73.30	93.04	79.16
PROC MIXED AR(1)	10	33.00	36.70	30.20	29.66	34.62	26.68
	20	67.36	69.36	63.58	61.24	73.20	58.34
	30	87.30	89.84	84.20	81.84	92.84	80.64

Table 9. Power for Detecting an Effect Size of $2\sigma_S$ for Time Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	73.06	49.52	70.12	44.26	86.06	46.88
	20	99.58	99.54	99.48	98.64	99.98	99.26
	30	100	100	100	100	100	100
PROC MIXED CS	10	91.64	94.68	93.36	95.22	99.34	98.70
	20	99.90	99.92	99.96	99.96	100	99.98
	30	100	100	100	100	100	100
PROC MIXED TOEP	10	85.66	86.56	83.98	79.84	94.02	83.62
	20	99.86	99.74	99.76	99.60	100	99.82
	30	100	100	100	100	100	100
PROC MIXED UN	10	73.06	49.52	70.12	44.26	86.06	46.88
	20	99.58	99.54	99.48	98.64	99.98	99.26
	30	100	100	100	100	100	100
PROC MIXED AR(1)	10	91.46	94.12	90.00	90.56	95.96	88.38
	20	99.88	99.94	99.92	99.88	100	99.86
	30	100	100	100	100	100	100

4.4. Power for Interaction Effect

Tables 10, 11 and 12 show the power results for detecting various effect sizes on the interaction effect. The trends observed from these powers for the interaction effect are like those observed for the power of detecting an effect size on the time effect. As sample size increased the power also increased. This trend can be visualized on Figure 12, which displays the PROC MIXED power analysis results for the various covariance structures when $k = 4$, $\rho = 0.75$, and effect size = σ_S . Figure 12 shows that compound symmetry resulted in the highest power. The same increasing trend in power was also true when the number of repeated measures increased. The trend by number of repeated measures can be visualized on Figure 13 for the PROC MIXED AR(1) power analyses when $k = 4$ and effect size = σ_S .

Once again, the trend for power with various correlations is different depending on the method used or the covariance structure specified. CS is again the only covariance structure

where an increase in correlation always resulted in an increase of power. The trends for power with each correlation can be seen in Figure 14, which displays the power analysis results from PROC MIXED AR(1) when $k = 4$ and effect size = σ_S . Figure 14 shows that the power is higher when correlation is equal to 0 and 0.75, than it is when correlation is equal to 0.25.

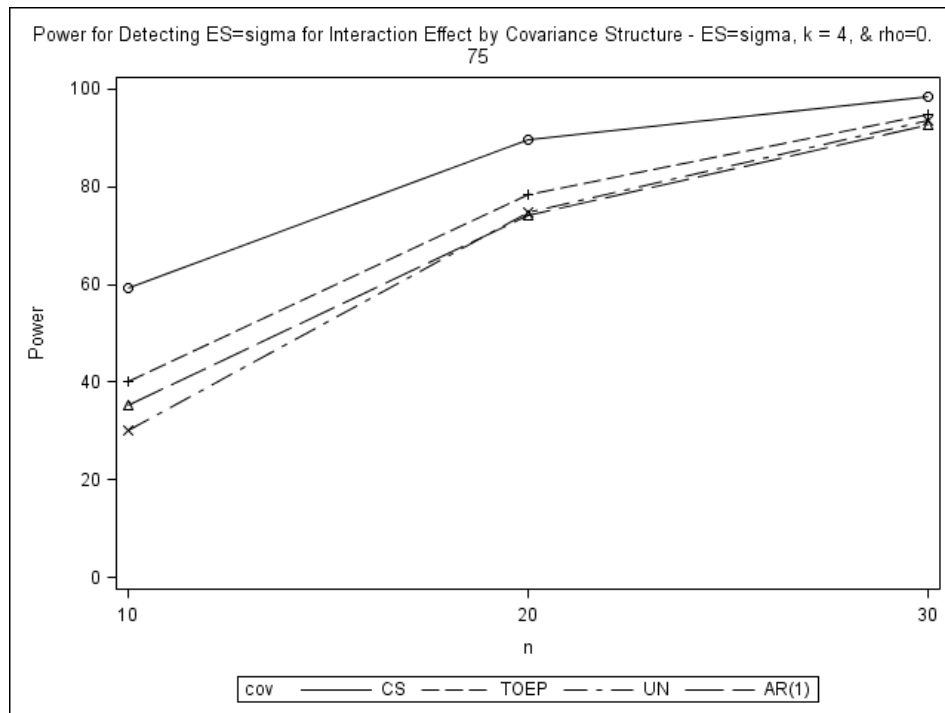


Figure 12. Interaction Effect – Trend of Power by Covariance Structure

The power results for an effect size of $(\frac{1}{2})\sigma_S$ on the interaction effect is on Table 10.

These results are very similar to those for the power analysis of the time effect when effect size was equal to $(\frac{1}{2})\sigma_S$. No combination of simulation parameters enables for the detection of $(\frac{1}{2})\sigma_S$ effect size on the interaction effect. Therefore, a sample size of greater than 30 would be needed for adequate power.

The power results for an effect size of σ_S on the interaction effect is on Table 11. To detect an effect size of σ on the interaction effect with a power of 90 percent, a sample size of at least 30 with 4 repeated measures, and a correlation of 0.75 was needed. Thus, to detect an effect

size of σ_S or less for the interactions effect with 90 percent power, a sample size of 30 or more is required.

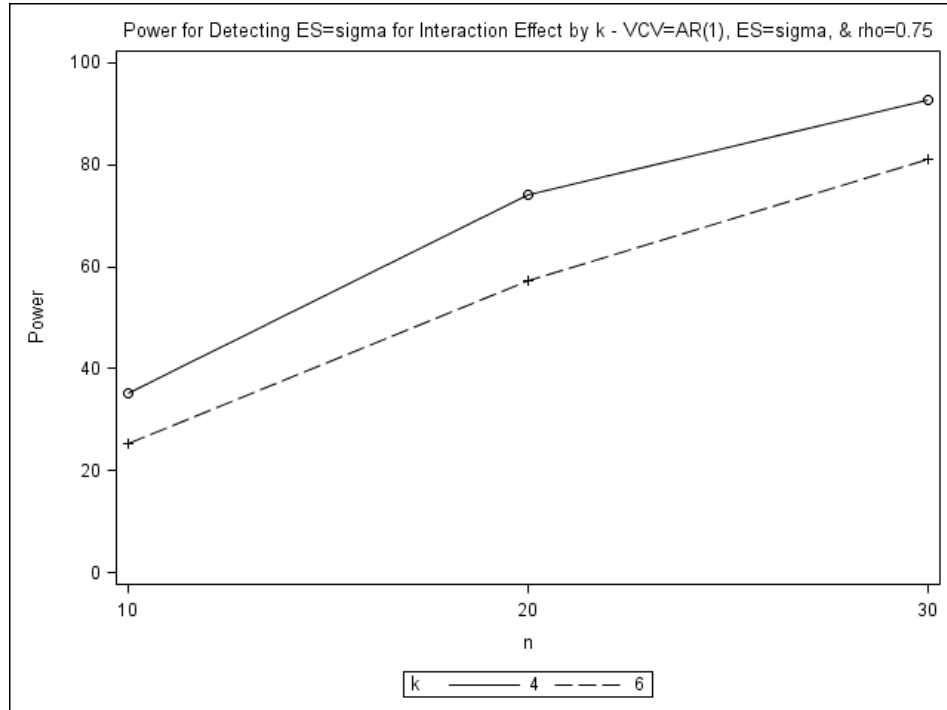


Figure 13. Interaction Effect – Trend of Power by Number of Repeated Measures

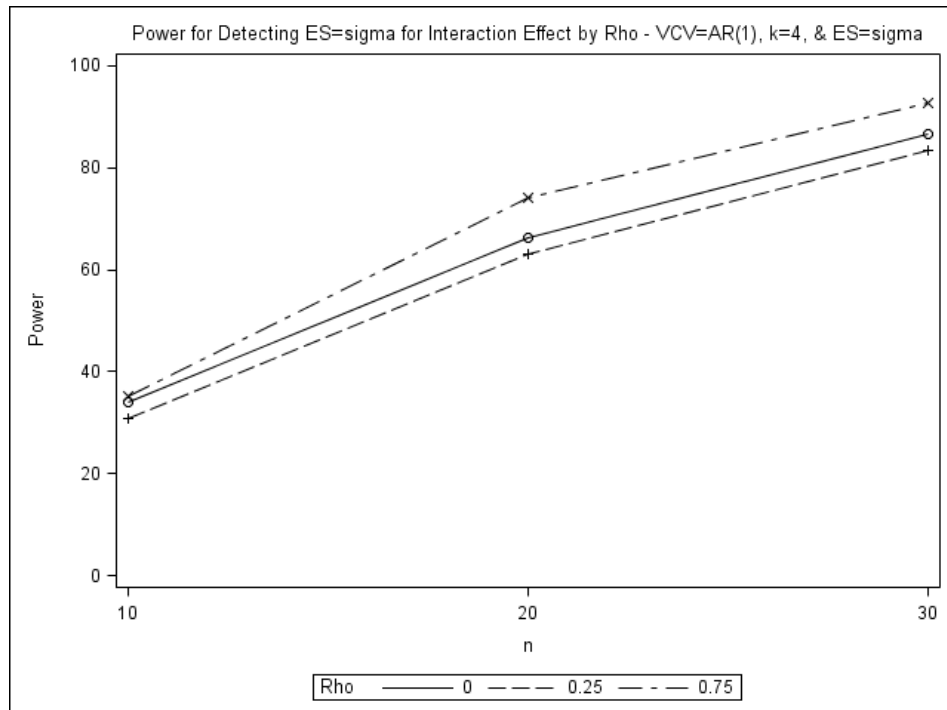


Figure 14. Interaction Effect – Trend of Power by Correlation

The power results for an effect size of $2\sigma_S$ on the interaction effect is on Table 12. To detect an effect size of $2\sigma_S$ for the interaction effect with 90 percent power, a sample size of at least 20, no matter the method or approach used in this study. The power by effect size can be seen on Figure 15. The powers plotted on Figure 15 are based in the PROC MIXED AR(1) power analyses when $k = 4$ and $\rho = 0.75$. The Type I errors for the treatment and time effect stayed controlled when simulating the interaction effect sizes. Table B13, B14, B15, B16, B17, and B18 in Appendix B show these Type I error rates.

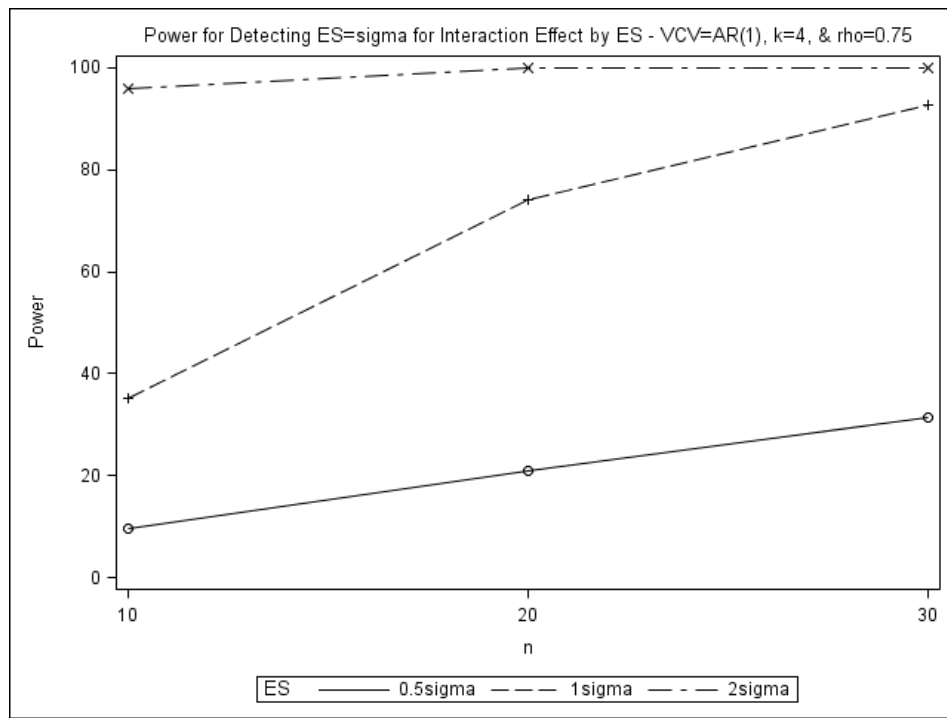


Figure 15. Interaction Effect – Trend of Power by Effect Size

Table 10. Power for Detecting an Effect Size of $(\frac{1}{2})\sigma_S$ for Interaction Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM	10	8.88	7.54	8.92	7.12	10.60	6.76
Wilk's	20	17.98	16.20	16.46	13.82	23.14	14.80
Lambda	30	25.68	24.54	25.40	20.44	35.72	23.56
PROC MIXED	10	11.60	11.16	12.16	12.84	18.94	18.80
	20	20.80	19.88	21.42	22.28	36.68	32.96
CS	30	28.98	29.96	32.12	31.20	51.06	46.06
PROC MIXED	10	10.62	11.36	10.56	10.82	12.96	10.58
	20	20.04	19.12	17.96	16.72	25.42	8.12
TOEP	30	28.12	29.28	26.98	23.50	37.88	25.96
PROC MIXED	10	8.88	7.54	8.92	7.12	10.60	6.78
	20	17.98	16.20	16.46	13.82	23.14	14.80
UN	30	25.68	24.54	25.40	20.44	35.72	23.56
PROC MIXED	10	10.84	10.76	10.30	10.26	9.66	7.50
	20	19.92	19.72	17.76	16.90	20.86	14.68
AR(1)	30	28.76	29.92	26.86	24.28	31.34	21.60

Table 11. Power for Detecting an Effect Size of σ_S for Interaction Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM	10	23.84	15.66	21.48	13.74	30.22	15.18
Wilk's	20	58.40	55.82	54.68	45.80	74.72	51.58
Lambda	30	82.36	82.24	79.52	73.80	93.60	78.82
PROC MIXED	10	34.66	36.40	37.34	38.68	59.24	54.04
	20	66.48	71.96	69.48	71.14	89.76	85.36
CS	30	86.60	89.46	87.46	88.78	98.40	96.22
PROC MIXED	10	32.38	31.72	28.88	27.20	40.02	27.96
	20	63.50	67.40	60.46	58.08	78.30	61.04
TOEP	30	85.22	87.20	82.30	80.42	94.88	84.20
PROC MIXED	10	23.84	15.66	21.48	13.74	30.22	15.18
	20	58.40	55.82	54.68	45.80	74.72	51.58
UN	30	82.36	82.24	79.52	73.80	93.60	78.82
PROC MIXED	10	34.06	36.38	30.84	29.56	35.32	25.30
	20	66.40	71.30	62.96	61.58	74.04	57.24
AR(1)	30	86.70	89.62	83.54	83.02	92.60	81.10

Table 12. Power for Detecting an Effect Size of $2\sigma_S$ for Interaction Effect

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM	10	71.92	50.02	69.88	43.78	85.44	47.00
Wilk's	20	99.70	99.64	99.54	98.64	100	99.14
Lambda	30	100	100	100	100	100	99.98
PROC	10	92.06	94.62	93.64	94.50	99.44	98.72
MIXED	20	99.92	100	99.94	99.98	100	100
CS	30	100	100	100	100	100	100
PROC	10	85.60	86.30	84.16	79.72	93.86	83.16
MIXED	20	99.90	99.94	99.70	99.76	99.98	99.80
TOEP	30	100	100	100	100	100	100
PROC	10	71.92	50.02	69.88	43.76	85.44	47.00
MIXED	20	99.70	99.64	99.54	98.64	100	99.14
UN	30	100	100	100	100	100	99.98
PROC	10	91.56	94.56	90.08	89.28	96.06	88.12
MIXED	20	99.92	100	99.92	99.94	100	99.88
AR(1)	30	100	100	100	100	100	100

CHAPTER 5. MOTIVATING EXAMPLE

To illustrate the findings from this study, a real dataset was analyzed. The dataset is from a study performed at NDSU in the Department of Health, Nutrition, and Exercise Sciences by Pulkrabek, Rhee, Gibbs, and Hall (2017). The goal of the study was to determine if diets containing different amounts of flax seed and/or buckwheat increased the amount of Enterobacteriaceae in the gastrointestinal tract. There was a total of 6 different diets that were tested. Each diet group had 12 mice as their experimental units from which multiple response variables were measured each week for eight weeks. The weight of each mouse was one of the response variables measured, and was the only response variable considered in this example.

Since the simulation study conducted for this paper only considered studies with two treatments, only two of the six diet groups from the weight data set were considered. This allows the data obtained by Pulkrabek et al. (2017) to better compare to the results of this simulation study. Therefore, the example data was considered to have only 2 treatment groups, 12 subjects per group, and 8 repeated measurements on each subject. The goal when analyzing this data was to determine the covariance structure that fit the data the best, and then compare how the results varied when testing the data with different covariance structures.

As suggested in Chapter 5 of SAS for Mixed Models, 2nd Edition (Littell et al, 2006a), and by Kincaid (2005), the example mice data was first run using PROC MIXED with the covariance structure specified as unstructured. The code for this step is shown in Figure 16, and output for the estimated covariance matrix and correlation matrix are shown on Figures 17 and 18 respectively.

After this information was obtained, the covariance estimates were plotted over the lag between the observed repeated variable, week, using the code provided by Littell et al. (2006a).

This plot, shown on Figure 19, allowed for visualization of the trends for the variances and covariances over time to better understand what covariance structure would fit the data the best.

This technique is again followed by the suggestions by Littell et al. (2006a).

```
proc mixed data=mice_uv;
  class trt week subj_id;
  model weight = trt|week;
  repeated week/subject =subj_id(trt)type=UN r rcorr;
run;
```

Figure 16. SAS Code to Obtain the Estimated Correlation Matrix using Unstructured

Estimated R Matrix for subj_id(trt) 13 Brown								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	2.2184	2.3001	2.9722	3.4098	4.0229	4.3822	4.5495	4.0846
2	2.3001	2.8428	3.6819	4.1363	4.9512	5.3188	5.4843	5.2513
3	2.9722	3.6819	5.1707	5.6398	6.8465	7.3003	7.5932	6.8235
4	3.4098	4.1363	5.6398	6.6784	7.9721	8.5120	8.7718	7.8149
5	4.0229	4.9512	6.8465	7.9721	9.8426	10.4549	10.7016	9.5181
6	4.3822	5.3188	7.3003	8.5120	10.4549	11.1860	11.4341	10.0873
7	4.5495	5.4843	7.5932	8.7718	10.7016	11.4341	11.9635	10.6743
8	4.0846	5.2513	6.8235	7.8149	9.5181	10.0873	10.6743	12.9526

Figure 17. SAS Output for the Estimated Covariance Matrix using Unstructured

Estimated R Correlation Matrix for subj_id(trt) 13 Brown								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.9159	0.8776	0.8859	0.8609	0.8797	0.8831	0.7620
2	0.9159	1.0000	0.9604	0.9493	0.9360	0.9432	0.9404	0.8654
3	0.8776	0.9604	1.0000	0.9597	0.9597	0.9599	0.9654	0.8338
4	0.8859	0.9493	0.9597	1.0000	0.9833	0.9848	0.9813	0.8402
5	0.8609	0.9360	0.9597	0.9833	1.0000	0.9964	0.9862	0.8430
6	0.8797	0.9432	0.9599	0.9848	0.9964	1.0000	0.9884	0.8380
7	0.8831	0.9404	0.9654	0.9813	0.9862	0.9884	1.0000	0.8575
8	0.7620	0.8654	0.8338	0.8402	0.8430	0.8380	0.8575	1.0000

Figure 18. SAS Output for the Estimated Correlation Matrix using Unstructured

The correlation and covariance matrices showed that unstructured was most likely the best fit for this data. One reason for this was because the variances differ for each week. This is shown in the covariance matrix in Figure 17 because the variances are along the diagonal get larger as the week increases. Usually in this case a heterogenous covariance structure would be

considered because they allow for different variances. Heterogenous covariance structures also require more parameter estimates than their homogenous version (eg. AR(1) compared to ARH(1)), but still have fewer than unstructured. For the purposes of this study only the covariance structures that were focused on in Chapter 4 will be considered so that this example stays within the scope of this paper. Another reason that unstructured was thought to possibly be the best fit was that the correlations did not follow a pattern like the ones required for AR(1) and TOEP, nor were all the correlations equal like a CS covariance structure requires. There did seem to be a possible gradual decline in the correlation matrix like for AR(1) or TOEP, but the covariance matrix showed a different trend where each covariance was different, which fits the unstructured covariance structure.

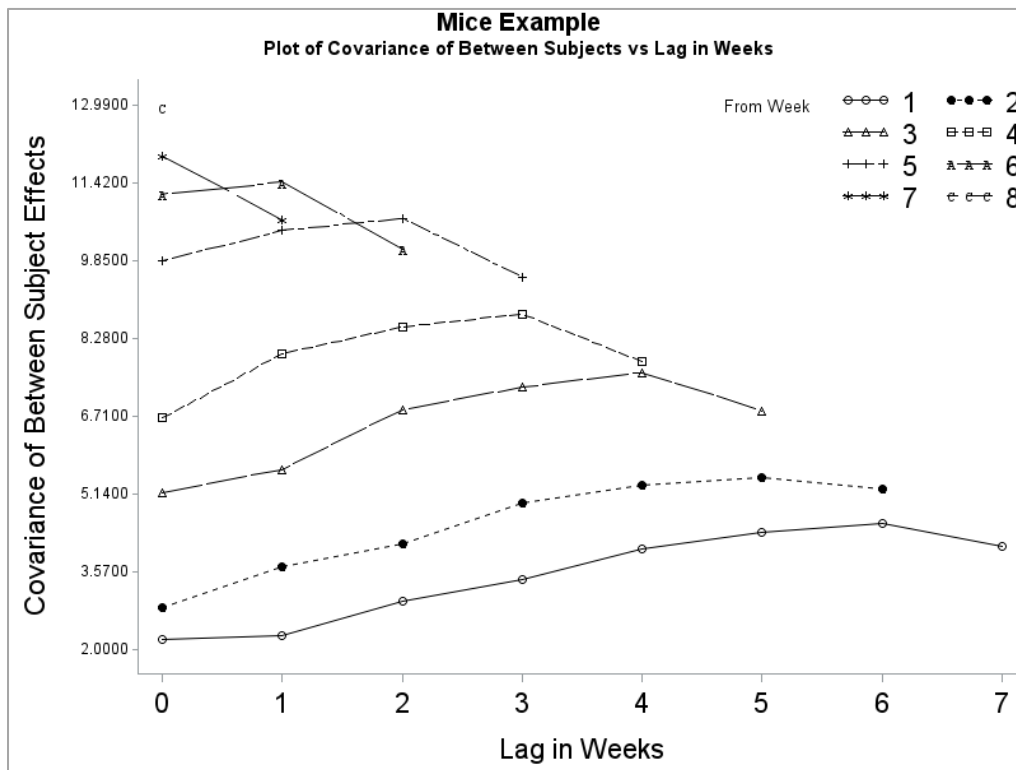


Figure 19. Plot of Covariance of Between-Subjects vs Lag in Weeks for UN

To confirm that UN was the correct covariance structure, plots of the covariance of between-subjects vs. the lag in weeks were plotted with the remaining three covariance

structures. These plots are shown on Figures 20, 21, and 22, and were compared to the plot on Figure 19. The trend on Figure 19 showed that all plotted lags from each week had a different variance, which is plotted at lag 0. Also, the covariances plotted from each week increased as the lag increased, except for a decrease in the covariance at the last lag in weeks plotted. The plots on Figure 20, 21, and 22 all showed overlapping lines which means that each week started at the same variance at lag 0 (indicating equal variances for each week), and had the same covariance values as lag increased. These plots also showed that the covariance decreased or stayed flat as lag increased, but this was not the trend in the unstructured plot on Figure 19. Therefore, it was observed that none of these plots fit the same trend shown by the unstructured plot in Figure 19, which again indicated that UN was the correct covariance structure.

Additionally, the fit statistics were compared to determine which covariance structure the data most likely had. Fit statistics are another good source to help determine which covariance structure best fits the data, and are computed by accounting for the residual estimates and number of parameters needed to be estimated. In SAS, the smaller the fit statistic, the better the fit. Note that this was also done for a simulation run in this study and the results can be seen on Tables B19, B20, and B21 in Appendix B. The results for the fit statistics for this example can be seen on Table 13. From these results, it was determined that the covariance structure with the lowest fit statistics was UN.

Table 13: Fit Statistics for Different Covariance Structures

Fit Statistic	AIC (Smaller is Better)	AICC (Smaller is Better)	BIC (Smaller is Better)
TOEP	571.1	571.9	580.5
UN	468.4	487.6	510.8
CS	666.7	666.8	669
AR	565.3	565.4	567.7

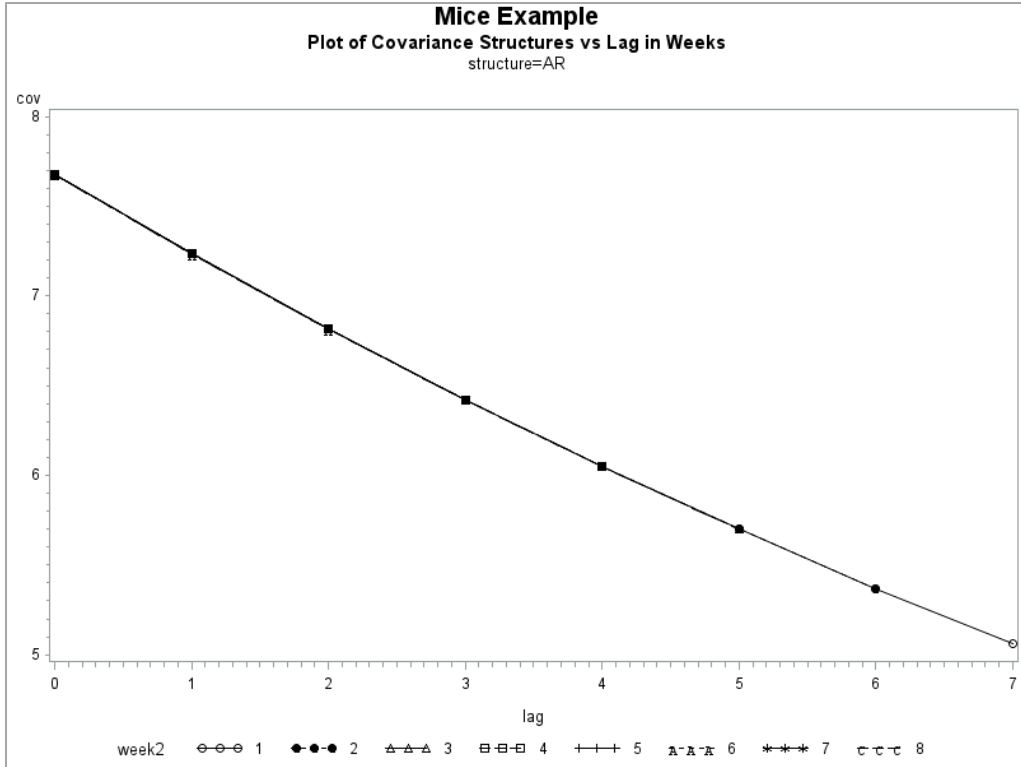


Figure 20. Plot of Covariance of Between-Subjects vs Lag in Weeks for AR(1)

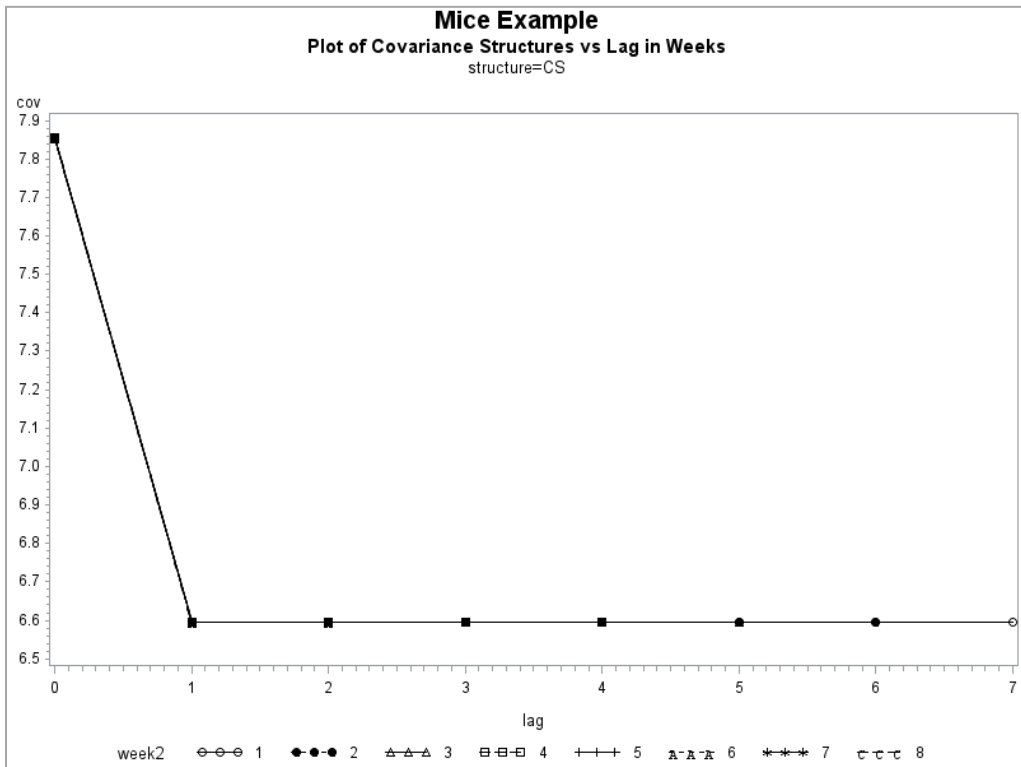


Figure 21. Plot of Covariance of Between-Subjects vs Lag in Weeks for CS

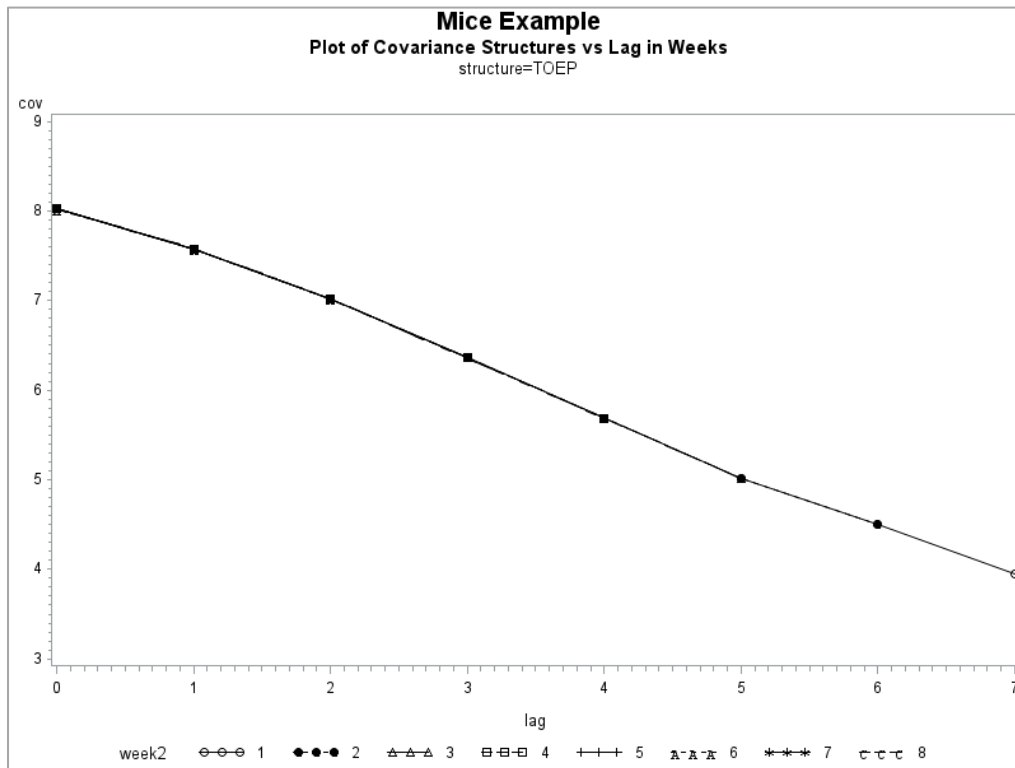


Figure 22. Plot of Covariance of Between-Subjects vs Lag in Weeks for TOEP

Since both the plots and fits statistics results had matching conclusions, UN was concluded to be the correct covariance structure for this data set. Therefore, how the p-values for the treatment, week, and interaction effects compared to the UN results was of interest. The p-value results are shown on Table 14.

It is shown on Table 14, that the p-values for the treatment effect are all around the same. This may go along with the results presented in Chapter 4 of this paper, because the power for the between-subjects factor did not show much variation with different covariance structures. The p-values for the week effect are all less than 0.0001, so all covariance structures yield the same conclusion. The power for the within-subjects factor in Chapter 4 of this paper showed that CS was the most powerful covariance structure, but that was with AR(1) being the true covariance structure. In this case, with the true covariance structure assumed to be UN, it is hard

to make a direct comparison. The p-values for the interaction term are the best for the unstructured covariance structure. Had the TOEP or AR(1) structure been chosen in this case, the interaction effect would have been concluded to be insignificant. It is interesting that the compound symmetry had the second smallest p-value, but this could be due to the simplicity of the structure. Overall, this example shows that choosing the incorrect covariance structure can impact the conclusions made on the effects being tested.

Table 14. P-values by Covariance Structure

Obs	Effect	NumDF	DenDF	FValue	ProbF	Structure
1	trt	1	25.3	1.07	0.3106	TOEP
2	trt	1	22	1.04	0.3191	UN
3	trt	1	22	1.04	0.3191	CS
4	trt	1	23.9	1.07	0.3106	AR
5	trt*week	7	62.5	1.47	0.1938	TOEP
6	trt*week	7	16	4.80	0.0045	UN
7	trt*week	7	154	2.61	0.0142	CS
8	trt*week	7	154	1.49	0.1732	AR
9	week	7	62.5	77.51	<.0001	TOEP
10	week	7	16	117.68	<.0001	UN
11	week	7	154	343.83	<.0001	CS
12	week	7	154	106.45	<.0001	AR

CHAPTER 6. CONCLUSION

There are many disciplines that use repeated measures when conducting their studies. As discussed in Chapter 2, most research on repeated measures is related to Type I error rate, how to conduct various methods of analysis, as well as determining the correct covariance structure. This study focused on the power analysis on repeated measures data with the objective to answer three main questions.

Question one was how does the performance of multivariate analysis compare to the performance of the mixed model method? The answer to this question can be based off the results discussed in Chapter 4. With respect to the between-subjects effect, treatment, the multivariate method with PROC GLM had reasonable power results. The PROC MIXED power analysis with AR(1) resulted in the highest power, which was expected because it was the true covariance structure. The multivariate method did not perform better than the power analysis with the true covariance structure specified, but did perform just as well as the mixed models approach when specifying the incorrect covariance structure.

With respect to the within-subjects effect, time, the multivariate method may not be preferred. When trying to detect a small effect size, the multivariate method seemed to perform similarly to the mixed model method. Although, when the effect size was large, the power for the multivariate method did not increase as much as the power did for some of the covariance structures specified for the mixed model. For example, when the effect size was 2σ , $\rho = 0.75$, $n = 10$, and $k = 6$, the multivariate method only had a power of 47 percent, while AR(1), CS, and TOEP all had a power of over 80 percent. This could be due to the multivariate method assuming an unstructured covariance structure. An unstructured covariance structure is more complex than the true covariance structure for the data, and therefore could be causing the low power. Overall,

if testing for the effect on the within-subjects effect is of interest, the multivariate method may not be the best approach if the true data has a simpler covariance structure. Note that the multivariate results were the same as the results from the mixed model method when UN was the specified covariance structure, and therefore choosing a covariance structure other than UN when interested in the within-subjects effect may also be preferred when the data truly has a simpler covariance structure and the sample size is small.

Question two was how does choosing the incorrect covariance structure affect the power? As just discussed, if researchers are only concerned about the between-subjects effect, then choosing the incorrect covariance structure may only have a slight impact on power. Choosing the correct covariance structure in this simulation study did result in the highest power, but if a researcher can get close to the correct covariance structure then the power should not be greatly affected. Choosing the incorrect covariance structure when testing for an effect on the within-subjects effect may be more important. The power analysis for the correct covariance structure, PROC MIXED with AR(1) specified as the covariance structure, did not result in the highest power for detecting an effect size in the within-subjects effect. The covariance structure with the lowest power was the PROC MIXED power analysis with UN as the specified covariance structure. The covariance structure with the highest power was the PROC MIXED power analysis with CS as the specified covariance structure. This is not surprising as unstructured is a complex covariance structure, so a low power is expected when sample sizes are small. Compound symmetry is one of the simplest covariance structures. Therefore, if not sure which error structure to choose, going with the slightly simpler choice may give a researcher more power to find a difference in the within-subjects effect.

Question three was what sample sizes are needed to produce adequate power of at least 90 percent for datasets with different simulation parameters? The answer to this question depends on the effect size being tested, and the simulation parameters being used. When testing for a difference in the between-subjects effect, and the ratio of the subject variance component to the residual variance component is high like in this study, researchers should keep in mind that a higher sample size and a higher number of repeated measures will result in the highest power. Also, the smaller the correlation, the higher the power. Therefore, to obtain adequate power of 90 percent for an effect size of σ , a study may only need 4 repeated measures and 20 total subjects if the correlation is small. Based on the simulation results in this study, the power trends for the within-subjects effect and the interaction effect are slightly different than the power trends for the between-subjects effect. Power still increases as sample size increases, but unlike the trend for the between-subjects effect, the power decreases as the number of repeated measures increases. This could be due to way the effect size was specified when generating the data sets as mentioned in the results section, and therefore may be a result due to the way this simulation study was conducted. Additionally, the results of this study indicate that for a small number of repeated measures, a correlation close to one results in a high power. Therefore, to obtain a power of at least 90 percent for detecting an effect size of σ for either the within-subjects effect, or the interaction effect, a researcher may need 4 repeated measures, a sample size of 30, and a high correlation.

Overall, the findings support that choosing the correlation that best fits the data will result in adequate power when sample sizes are high. Choosing the slightly simpler covariance structure may be preferred over choosing the more complex covariance structures because complex covariance structures have more parameters to estimate and can therefore reduce power.

Additionally, one hope of this study is that researchers can use the tables presented in the results section as an informal guide to determine the sample size needed for adequate power of a future study. By looking at these tables of power and considering the simulation parameters that best fit the data being used and the effect size that is needing to be detected, recommendations on sample size and number of repeated measures could be made.

The research conducted in this paper could be continued on in many ways to obtain more knowledge about the power of repeated measures. Simulating data with different ratios of subject variance components to residual variance components and then determining power trends would be a good continuation of this study. Additionally, conducting a power analysis on repeated measures data that is unbalanced, or follows a non-normal distribution are also future research suggestions.

REFERENCES

- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical & Statistical Psychology*, 31. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8317.1978.tb00581.x/abstract>
- Everitt, B.S. (1995). The Analysis of Repeated Measures: A Practical Review with Examples. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 44 (1). Retrieved from <http://www.jstor.org/stable/2348622>
- Field, A. (2010). A Bluffer's Guide to ... Sphericity. *BPS-MSc Newsletter 6 (1)*. Retrieved from <http://www.discoveringstatistics.com/docs/sphericity.pdf>
- Gueorguieva, R. & Krystal, J. H. (2004). Move Over ANOVA: Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. *Archives of general psychiatry*, 61. Retrieved from https://www.researchgate.net/profile/Ralitza_Gueorguieva/publication/5412094_Move_Over_ANOVA_Progress_in_Analyzing_Repeated-Measures_Data_andIts_Reflection_in_Papers_Published_in_the_Archives_of_General_Psychiatry/links/5617793308ae40a7199a94ca/Move-Over-ANOVA-Progress-in-Analyzing-Repeated-Measures-Data-andIts-Reflection-in-Papers-Published-in-the-Archives-of-General-Psychiatry.pdf
- Guerin, L. & Stroup, W. W. (2000). A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data. *Annual Conference on Applied Statistics in Agriculture*. Retrieved from <http://newprairiepress.org/agstatconference/2000/proceedings/15>

- Kenward, M. G. & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53(3). Retrieved from <http://www.public.iastate.edu/~dnett/S511/KR.pdf>
- Keselman, H. J., Algina, J., Kowalchuk, R. K. & Wolfinger, R. D. (1999). The Analysis of Repeated Measurements: A Comparison of Mixed-Model Satterthwaite F Tests and a Nonpooled Adjusted Degrees of Freedom Multivariate Test. *Communications in Statistics - Theory and Methods*, 28(12). Retrieved from <http://home.cc.umanitoba.ca/~kesel/cis1999b.pdf>
- Kiernan, K., Tao, J., & Gibbs, P. (2012). Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures. *SAS Global Forum 2012*. Retrieved from <http://support.sas.com/resources/papers/proceedings12/332-2012.pdf>
- Kincaid, C. (2005). Guidelines for Selecting the Covariance Structure in Mixed Model Analysis. *SUGI 30 Proceedings*. Retrieved from <http://www2.sas.com/proceedings/sugi30/198-30.pdf>
- Kowalchuk, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The Analysis of Repeated Measurements with Mixed-Model Adjusted F Tests. *Educational and Psychological Measurement*, 64. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/0013164403260196>
- Littell, R. C. (2011). The Evolution of Linear Models in SAS: A Personal Perspective. *SAS Global Forum 2011*. Retrieved from <http://support.sas.com/resources/papers/proceedings11/325-2011.pdf>
- Littell, R.C., Henry, P.R., & Ammerman, C.J. (1998). Statistical Analysis of Repeated Measures Data Using SAS Procedures. *J. Animal Science*, 76. Retrieved from

http://www.stat.ncsu.edu/people/arellano/courses/ST524/Fall08/Homeworks/Homework7/articles/Littell_mixed_JAS.pdf

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006a).

Chapter 5: Analysis of Repeated Measures Data. *SAS® for Mixed Models, Second Edition* (pp. 160-186). Cary, NC: SAS Institute.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006b).

Appendix 1: Linear Mixed Model Theory. *SAS® for Mixed Models, Second Edition* (pp. 734-756). Cary, NC: SAS Institute.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the

analysis of repeated measures data. *Statistics in Medicine*, *19*(13). Retrieved from

[http://onlinelibrary.wiley.com/doi/10.1002/1097-0258\(20000715\)19:13%3C1793::AID-SIM482%3E3.0.CO;2-Q/abstract](http://onlinelibrary.wiley.com/doi/10.1002/1097-0258(20000715)19:13%3C1793::AID-SIM482%3E3.0.CO;2-Q/abstract)

Moser, E. B. (2004). Repeated Measures Modeling With PROC MIXED. *SUGI 29 Proceedings*.

Retrieved from <http://www2.sas.com/proceedings/sugi29/188-29.pdf>

Oberfeld, D. & Franke, T. (2013) Evaluating the robustness of repeated measures analyses: The

case of small sample sizes and nonnormal data. *Behav Res*, *45*. Retrieved from

https://www.staff.uni-mainz.de/oberfeld/downloads/oberfeld_franke_2013_BRM.pdf

Pulkrabek, M., Rhee, Y., Gibbs, P., & Hall, C. (2017). Flaxseed- and Buckwheat-Supplemented

Diets Altered Enterobacteriaceae Diversity and Prevalence in the Cecum and Feces of

Obese Mice. *Journal of Dietary Supplements*, *14*(6). Retrieved from

<http://www.tandfonline.com/doi/abs/10.1080/19390211.2017.1305477?journalCode=ijds>

SAS Institute Inc. (2017). *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

Retrieved from <http://support.sas.com/documentation/onlinedoc/stat/143/statug.pdf>

Stroup, W. W. (1999). On Using PROC MIXED for Longitudinal Data. *Annual Conference on Applied Statistics in Agriculture*. Retrieved from

<http://newprairiepress.org/agstatconference/1999/proceedings/5>

Wang, Z. & Goonewardene, L. A. (2004). The use of MIXED models in the analysis of animal experiments with repeated measures data. *Canadian Journal of Animal Science*.

Retrieved from [https://era.library.ualberta.ca/files/crb68xb84k/cjas_84\(1\)](https://era.library.ualberta.ca/files/crb68xb84k/cjas_84(1))

Wolfinger, R. D. (1996). Heterogeneous Variance: Covariance Structures for Repeated Measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1(2).

Retrieved from

https://www.researchgate.net/profile/Russ_Wolfinger2/publication/272579574_Heterogeneous_Variance_Covariance_Structures_for_Repeated_Measures/links/5756da6808ae5c654903f3fe/Heterogeneous-Variance-Covariance-Structures-for-Repeated-Measures.pdf

Wolfinger, R. D., & Chang, M. (1995). Comparing the SAS GLM and MIXED Procedures for Repeated Measures. *SUGI Proceedings*. Retrieved from

<https://support.sas.com/rnd/app/stat/papers/mixedglm.pdf>

APPENDIX A. SIMULATION CODE

The code below was used to perform the simulation study. Various parts of the code were changed to obtain all the simulations, such as the parameters k , s , ρ , and the effect size specified in β vector. Note that some of the code is commented off to prevent large amounts of output from printing.

```
proc iml;

  %let seed0=0;
  %let NumSamples=5000;
  t=2;          *** t defines # of treatment groups ***;
  k=4;          *** k defines # of repeated measures ***;
  s=10;         *** s defines # of subjects ***;
  rho=0.25;     *** Autocorrelation for AR(1) ***;
  sigma2_R=1;   *** Residual variance component ***;
  sigma2_S=4;   *** Subject variance component ***;
  NumSamples = &NumSamples;

  ***** VCV basic structure is AR(1) here *****;
  B=j(k,k,0);   *** Matrix of zeroes. ***;
  do i=1 to k;
    do j=1 to k;
      if i=j then B[i,j]=1; *** Convert diagonal to 1s ***;
      else B[i,j]=rho**abs(i-j); *** AR(1) rhos ***;
    end;
  end;
  *print B;

  B = sigma2_S*B; *** Multiply VCV by subject sigma2 ***;
  *print B;
  R = I(s) @ B;   *** block-diagonal matrix ***;
  *print R;

  ***** Set up the fixed portion design *****;
  Trt=j((s/t*k),1,1); *** Set up trt design columns ***;
  do i=2 to t;
    Trt=colvec(Trt//j((s/t*k),1,i));
  end;
  Trt = design(Trt);
  *print Trt;

  xxx=I(k);      *** Set up repeated measures columns ***;
  do i=1 to s;
    Week=(Week//xxx);
  end;

  X = j(nrow(Week),1,1)||Trt||Week;
  *print X;

  ***** Subject Random Error *****;
```

```

call randseed(&seed0);
zero = j(1, k*s, 0);          *** the zero vector ***;
print zero;
esubj = RandNormal(NumSamples, zero, R); * esubj ~ MVN(0,R);
esubject=t(esubj);
*print esubject;

***** Fixed effects parms: mu t1 t2 w1 w2 w3 w4 *****;
beta = {5, 0, 0, 0, 0, 0, 0}; *** Can specify effect size in beta ***;
etal = X*beta;          *** Fixed effects portion of obs. ***;
eta = repeat(etal,1,NumSamples);
print beta;
*print eta;

***** Random Error Component -- eps[i,j]~N(0,sigma) *****;
eps = J(nrow(X), NumSamples); *** Residual error term ***;
call randgen(eps, "Normal", 0, sqrt(sigma2_R));
*print eps;

Y = eta + esubject + eps;
*print Y;

Indiv = colvec(repeat(T(1:s),1,k)); *** Create Subj_ID ***;
print Indiv;

MVdata=X[,2:(t+k+1)]||Indiv||Y;
*print mvdata;

create Subj_MV from MVdata;
append from MVdata;
close;

quit;

proc print data=Subj_MV(obs=2);          *** Print to check the data ***
title2 'MV Data Structure';
run;

data Subj_UV(keep=Sample Subj_ID Stress Trt Time);
set Subj_MV;
Subj_ID=Col7;
if col1=1 then trt='A';
else if col1=0 then trt='B';
if col3=1 then time=1;
else if col4=1 then time=2;
else if col5=1 then time=3;
else if col6=1 then time=4;
array ys{&NumSamples} col8-col15007; *** Adjust with sample size. ***;
do Sample=1 to &NumSamples;
Stress=ys{Sample};
output;
end;
run;

proc sort data=Subj_UV;

```

```

    by Sample;
    run;
proc print data=Subj_UV (obs=10);    *** Print to check the data ***
    title2 'UV Data Structure';
    run;

libname rm 'C:\Users\taylo\Documents\Thesis\';

%macro mvn(dsn,p,n);    *** dsn=actual VCV structure, p= # of time period,
                        n=sample size***;

proc transpose data=Subj_UV out=rm.mv_&dsn._&p._&n prefix=Stress;
    by Sample Subj_ID Trt;
    var Stress;
    run;
*proc print;
*run;

*****
*** Repeated Measures MV Analysis Block                                     ***
*****;
ods listing close;
ods output MultStat=MultStat_&dsn._&p
           ModelANOVA=ModelANOVAmv_&dsn._&p
           Epsilons=Epsilons_&dsn._&p;
proc glm data=rm.mv_&dsn._&p._&n;
    by sample;
    class trt;
    model stress1 - stress&p = trt / nouni;
    repeated time &p;
    title2 'Repeated Measures ANOVA for Effect of Time on Stress Level';
    run;
ods listing;

proc contents noprint;
    run;

data rm.uv_&dsn._&p._&n;
    set Subj_UV;
    run;

*****
*** Repeated Measures UV Analysis Block -- Traditional Approach           ***
*****;
ods listing close;
ods output ModelANOVA=ModelANOVAv_&dsn._&p
           FitStatistics=FitStatisticsuv_&dsn._&p;
proc glm data=rm.uv_&dsn._&p._&n;
    by sample;
    class trt time;
    model stress = trt | time;    *** allows subject error combined w/ mse.;
    title2 'Univariate ANOVA of Time on Stress Level (No Subject Effect)';
    run;
ods listing;

```

```

proc contents noprint;
  run;

*****
*** Repeated Measures UV Analysis Block -- Mixed Model Approach      ***
*****;
%macro mmuv (VCV);

*ods trace on / listing;
ods listing close;
ods output
  CovParms=Cov&VCV._&dsn._&p
  FitStatistics=Fit&VCV._&dsn._&p
  Tests3=Tests3&VCV._&dsn._&p;
proc mixed data=rm.uv_&dsn._&p._&n;
  by sample;
  class trt time subj_id;
  model stress = trt | time / ddfm=kr;
  repeated time / subject=subj_id type=&VCV;
  title2 "Repeated Measures ANOVA using Mixed Model Approach -- &VCV";
  run;
ods listing;

proc contents noprint;
  run;

%mend mmuv;

*** Calls to mmuv allow specifying the *****
*** covariance structure for a model fit ***;
%mmuv (VC);
%mmuv (CS);
%mmuv (TOEP);
%mmuv (UN);

ods listing close;
ods output
  CovParms=CovAR_&dsn._&p
  FitStatistics=FitAR_&dsn._&p
  Tests3=Tests3AR_&dsn._&p;
proc mixed data=rm.uv_&dsn._&p._&n;
  by sample;
  class trt time subj_id;
  model stress = trt | time / ddfm=kr;
  repeated time / subject=subj_id type=AR(1);
  title2 "Repeated Measures ANOVA using Mixed Model Approach -- AR(1)";
  run;
ods listing;

proc contents noprint;
  run;

title1 "Simulation of MV Normal Data, N=&N -- Data Covariance Structure was
&dsn";
proc sort data=MultStat_&dsn._&p (where=(Hypothesis='time' and
Statistic="Wilks' Lambda"))

```

```

        out=Per_mvrm(keep=Sample value fvalue probf);
    by sample;
    run;
proc sort data=MultStat_&dsn._&p (where=(Hypothesis='time_trt' and
Statistic="Wilks' Lambda"))
        out=PerTrt_mvrm(keep=Sample value fvalue probf);
    by sample;
    run;
/*
proc sort data=MultStat_&dsn._&p
        out=mvrm;
    by sample;
    run;
*/
proc sort data=ModelANOVAmv_&dsn._&p (where=(Source='time'))
        out=Per_mvghf(keep=Sample FValue ProbF ProbFGG ProbFHF);
    by sample;
    run;
proc sort data=ModelANOVAmv_&dsn._&p (where=(Source='time*trt'))
        out=PerTrt_mvghf(keep=Sample FValue ProbF ProbFGG ProbFHF);
    by sample;
    run;
proc sort data=ModelANOVAmv_&dsn._&p (where=(Source='trt'))
        out=Trt_mvghf(keep=Sample FValue ProbF ProbFGG ProbFHF);
    by sample;
    run;
/* ANOVA with GG and HF above */

proc sort data=ModelANOVAuv_&dsn._&p (where=(HypothesisType=3 and
source='time'))
        out=Per_uvrm(keep=Sample FValue ProbF SS MS);
    by sample;
    run;
proc sort data=ModelANOVAuv_&dsn._&p (where=(HypothesisType=3 and
source='trt*time'))
        out=PerTrt_uvrm(keep=Sample FValue ProbF SS MS);
    by sample;
    run;
proc sort data=ModelANOVAuv_&dsn._&p (where=(HypothesisType=3 and
source='trt'))
        out=Trt_uvrm(keep=Sample FValue ProbF SS MS);
    by sample;
    run;
/* ANOVA UV above */

proc sort data=Tests3VC_&dsn._&p (where=(Effect='time'))
        out=Per_uvrmVC(keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3VC_&dsn._&p (where=(Effect='trt*time'))
        out=PerTrt_uvrmVC(keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3VC_&dsn._&p (where=(Effect='trt'))
        out=Trt_uvrmVC(keep=Sample FValue ProbF);
    by sample;

```

```

    run;
  /*
proc sort data=Tests3VC_&dsn._&p
    out=uvrmVC;
    by sample;
    run;
  */

proc sort data=Tests3CS_&dsn._&p (where=(Effect='time'))
    out=Per_uvrmCS (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3CS_&dsn._&p (where=(Effect='trt*time'))
    out=PerTrt_uvrmCS (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3CS_&dsn._&p (where=(Effect='trt'))
    out=Trt_uvrmCS (keep=Sample FValue ProbF);
    by sample;
    run;
  /* CS Results above */

proc sort data=Tests3TOEP_&dsn._&p (where=(Effect='time'))
    out=Per_uvrmTOEP (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3TOEP_&dsn._&p (where=(Effect='trt*time'))
    out=PerTrt_uvrmTOEP (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3TOEP_&dsn._&p (where=(Effect='trt'))
    out=Trt_uvrmTOEP (keep=Sample FValue ProbF);
    by sample;
    run;
  /* TOEP Results above */

proc sort data=Tests3UN_&dsn._&p (where=(Effect='time'))
    out=Per_uvrmUN (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3UN_&dsn._&p (where=(Effect='trt*time'))
    out=PerTrt_uvrmUN (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3UN_&dsn._&p (where=(Effect='trt'))
    out=Trt_uvrmUN (keep=Sample FValue ProbF);
    by sample;
    run;
  /* UN Results above */

proc sort data=Tests3AR_&dsn._&p (where=(Effect='time'))
    out=Per_uvrmAR (keep=Sample FValue ProbF);
    by sample;
    run;
proc sort data=Tests3AR_&dsn._&p (where=(Effect='trt*time'))
    out=PerTrt_uvrmAR (keep=Sample FValue ProbF);

```

```

    by sample;
    run;
proc sort data=Tests3AR_&dsn._&p(where=(Effect='trt'))
    out=Trt_uvrmAR(keep=Sample FValue ProbF);
    by sample;
    run;
/* AR Results above */

/*
proc print data=mvghf; *** Print to check the data ***
    title2 'MVRM Structure';
    run;
*/

proc format;      *** make formats to reference later in proc freq ***;
    value pow low-.05 = 'Reject'
           .05<-high = 'DNR';
    value mod 1='Wilks'  2='UVCS'  3='GG'  4='HF'  5='UVNoBlk'
           6='MxVC'    7='MxCS'  8='MxTOEP'
           9='MxUN'   10='MxAR';
    value modtrt
           1='UVCS'  2='UVNoBlk'
           3='MxVC'  4='MxCS'  5='MxTOEP'
           6='MxUN'  7='MxAR';
    run;

*** Time tests are merged into one data set ***;
data rm.Per_all_ps_&p._&n;
    merge Per_mvrm (rename=(value=wilks fvalue=wl_f probf=wlf_p))
          Per_mvghf (rename=(fvalue=per_f probf=per_p))
          Per_uvrm (rename=(fvalue=uvper_f probf=uvper_p))
          Per_uvrmVC (rename=(fvalue=VCper_f probf=VCper_p))
          Per_uvrmCS (rename=(fvalue=CSper_f probf=CSper_p))
          Per_uvrmTOEP (rename=(fvalue=TOEPper_f probf=TOEPper_p))
          Per_uvrmUN (rename=(fvalue=UNper_f probf=UNper_p))
          Per_uvrmAR (rename=(fvalue=ARper_f probf=ARper_p));
*by sample;
run;
data Per_ult(keep=sample i p_value); *** renaming pvalues ***;
    set rm.Per_all_ps_&p._&n;
    array ps{10} wlf_p per_p ProbFGG ProbFHF uvper_p VCper_p CSper_p TOEPper_p
    UNper_p ARper_p;
    do i=1 to 10;
        p_value=ps{i};
        output;
    end;
run;
proc freq data=Per_ult; *** using the pervious format to calculate and label
the p-values ***;
    tables p_value*i / nopct norow;
    format p_value pow. i mod.;
    title2 'Time Results';
run;
*** Time*Trt Tests are merged into one data set ***;
data rm.PerTrt_all_ps_&p._&n;
    merge PerTrt_mvrm (rename=(value=wilks fvalue=wl_f probf=wlf_p))

```



```

    PerTrt_mvghf(rename=(fvalue=per_f probf=per_p))
    PerTrt_uvrn  (rename=(fvalue=uvper_f probf=uvper_p))
    PerTrt_uvrnVC(rename=(fvalue=VCper_f probf=VCper_p))
    PerTrt_uvrnCS(rename=(fvalue=CSper_f probf=CSper_p))
    PerTrt_uvrnTOEP(rename=(fvalue=TOEPper_f probf=TOEPper_p))
    PerTrt_uvrnUN(rename=(fvalue=UNper_f probf=UNper_p))
    PerTrt_uvrnAR(rename=(fvalue=ARper_f probf=ARper_p));

run;
data PerTrt_ult(keep=sample i p_value); *** renaming pvalues ***;
    set rm.PerTrt_all_ps_&p._&n;
    array ps{10} wlf_p per_p ProbFGG ProbFHF uvper_p VCper_p CSper_p TOEPper_p
UNper_p ARper_p;
    do i=1 to 10;
        p_value=ps{i};
        output;
    end;
run;
proc freq data=PerTrt_ult; *** using the pervious format to calculate and
lable the p-values ***;
    tables p_value*i / nopct norow;
    format p_value pow. i mod.;
    title2 'Time*Trt Results';
run;

*** Trt Tests are merged into one data set ***;
data rm.Trt_all_ps_&p._&n;
    merge Trt_mvghf(rename=(fvalue=per_f probf=per_p))
    Trt_uvrn  (rename=(fvalue=uvper_f probf=uvper_p))
    Trt_uvrnVC(rename=(fvalue=VCper_f probf=VCper_p))
    Trt_uvrnCS(rename=(fvalue=CSper_f probf=CSper_p))
    Trt_uvrnTOEP(rename=(fvalue=TOEPper_f probf=TOEPper_p))
    Trt_uvrnUN(rename=(fvalue=UNper_f probf=UNper_p))
    Trt_uvrnAR(rename=(fvalue=ARper_f probf=ARper_p));

run;
data Trt_ult(keep=sample i p_value); *** renaming pvalues ***;
    set rm.Trt_all_ps_&p._&n;
    array ps{7} per_p uvper_p VCper_p CSper_p TOEPper_p UNper_p ARper_p;
    do i=1 to 7;
        p_value=ps{i};
        output;
    end;
run;
proc freq data=Trt_ult; *** using the pervious format to calculate and lable
the p-values ***;
    tables p_value*i / nopct norow;
    format p_value pow. i modtrt.;
    title2 'Trt Results';
run;

%mend mvn;
%mvn(AR1, 4, 10);

%let t1 = %sysfunc(datetime());
%let elapsedTime = %sysevalf(&t1-&t0);
%put &elapsedTime;

```

APPENDIX B. TYPE I ERROR RATES AND FIT STATISTICS

Table B1. Type I Error for Time Effect when Effect Size on Treatment is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.92	5.62	5.04	5.06	5.30	4.96
	20	5.26	5.12	4.34	5.38	5.04	5.14
	30	5.38	5.28	4.74	4.90	5.04	4.12
PROC MIXED CS	10	4.66	5.50	4.90	4.82	5.78	6.32
	20	5.14	5.44	4.56	5.52	5.94	6.16
	30	5.22	5.14	4.68	5.08	5.34	5.70
PROC MIXED TOEP	10	5.28	6.24	5.46	5.72	5.44	5.24
	20	5.32	5.12	4.02	5.86	5.18	5.12
	30	5.18	5.40	4.44	4.80	5.04	4.10
PROC MIXED UN	10	4.92	5.62	5.04	5.08	5.30	4.96
	20	5.26	5.12	4.34	5.38	5.04	5.14
	30	5.38	5.28	4.74	4.90	5.04	4.12
PROC MIXED AR(1)	10	4.42	5.66	4.34	4.22	3.86	3.92
	20	5.04	5.30	4.16	5.28	4.42	4.34
	30	5.20	4.94	4.60	4.72	4.56	3.96

Table B2. Type I Error for Interaction Effect when Effect Size on Treatment is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.16	5.86	5.02	5.12	4.74	5.08
	20	5.08	4.44	4.90	4.98	5.50	4.90
	30	5.38	5.18	5.30	5.26	5.16	5.10
PROC MIXED CS	10	5.06	5.06	5.18	5.22	5.26	6.48
	20	4.96	5.02	5.18	5.12	6.12	6.34
	30	4.90	4.78	5.44	5.58	5.24	6.48
PROC MIXED TOEP	10	5.32	6.02	5.12	6.28	4.84	6.08
	20	4.86	5.26	4.82	5.04	5.62	4.76
	30	4.96	4.88	5.38	5.42	4.92	5.06
PROC MIXED UN	10	5.16	5.86	5.02	5.14	4.74	5.08
	20	5.08	4.44	4.90	4.98	5.50	4.90
	30	5.38	5.18	5.30	5.26	5.16	5.10
PROC MIXED AR(1)	10	4.96	5.36	4.40	4.28	3.66	3.78
	20	5.12	5.08	4.90	4.70	4.60	4.20
	30	4.90	5.02	4.80	5.06	4.04	4.18

Table B3. Type I Error for Time Effect when Effect Size on Treatment is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.46	4.94	4.80	5.70	4.94	5.26
	20	4.94	4.92	5.18	5.22	5.42	5.10
	30	4.68	5.40	5.16	5.46	5.36	5.08
PROC MIXED CS	10	4.94	5.56	5.36	5.54	5.46	6.72
	20	5.10	4.74	5.20	5.76	5.82	7.04
	30	4.84	6.08	5.20	5.58	5.54	7.00
PROC MIXED TOEP	10	5.16	5.62	5.24	5.96	5.06	6.30
	20	4.86	4.98	4.86	5.26	5.18	5.32
	30	4.60	5.96	4.98	5.26	5.32	5.60
PROC MIXED UN	10	5.46	4.94	4.80	5.70	4.94	5.26
	20	4.94	4.92	5.18	5.22	5.42	5.10
	30	4.68	5.40	5.16	5.46	5.36	5.08
PROC MIXED AR(1)	10	5.06	5.20	4.60	5.00	3.62	4.38
	20	4.78	4.78	4.70	5.44	4.22	4.28
	30	4.78	5.72	5.10	5.04	4.24	4.96

Table B4. Type I Error for Interaction Effect when Effect Size on Treatment is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.80	4.50	5.28	4.76	4.80	4.62
	20	5.02	5.58	5.06	5.36	5.32	5.08
	30	5.10	5.56	5.18	5.26	4.24	4.50
PROC MIXED CS	10	5.28	4.88	4.90	5.40	5.72	6.44
	20	4.90	5.28	5.26	5.58	6.10	6.06
	30	4.78	5.20	5.58	5.46	4.96	6.46
PROC MIXED TOEP	10	5.40	5.78	5.66	5.26	5.02	5.42
	20	5.02	5.36	5.04	5.32	5.72	4.62
	30	4.70	5.60	5.00	5.54	4.56	4.54
PROC MIXED UN	10	4.80	4.50	5.28	4.76	4.80	4.64
	20	5.02	5.58	5.06	5.36	5.32	5.08
	30	5.10	5.56	5.18	5.26	4.24	4.50
PROC MIXED AR(1)	10	5.14	5.00	4.72	4.68	4.18	3.64
	20	4.68	5.34	5.16	5.04	4.76	4.20
	30	4.74	5.36	4.98	5.14	4.04	4.26

Table B5. Type I Error for Time Effect when Effect Size on Treatment is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.98	5.12	4.48	5.08	4.82	4.94
	20	5.04	4.80	4.94	5.08	5.04	4.50
	30	5.08	5.26	5.18	4.68	4.92	5.58
PROC MIXED CS	10	4.94	5.20	5.70	5.50	4.70	6.24
	20	4.88	4.72	5.20	5.70	6.16	6.18
	30	5.10	4.90	5.48	5.38	5.86	6.94
PROC MIXED TOEP	10	5.48	5.48	5.84	6.10	4.76	5.54
	20	4.94	4.90	4.94	5.28	5.58	4.80
	30	4.92	5.14	5.32	4.94	5.06	5.18
PROC MIXED UN	10	4.98	5.12	4.48	5.08	4.82	4.94
	20	5.04	4.80	4.94	5.16	5.04	4.50
	30	5.08	5.26	5.18	4.68	4.92	5.58
PROC MIXED AR(1)	10	4.84	5.06	4.90	4.94	3.36	4.02
	20	4.98	4.68	4.86	5.06	4.64	4.00
	30	5.18	4.86	5.12	4.98	4.04	4.40

Table B6. Type I Error for Interaction Effect when Effect Size on Treatment is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.20	4.92	4.74	4.94	4.96	4.58
	20	4.90	4.86	5.46	5.56	4.80	5.50
	30	5.32	5.54	5.16	5.26	4.56	5.40
PROC MIXED CS	10	4.62	4.20	5.06	5.06	6.24	6.60
	20	4.64	5.06	5.76	5.66	5.02	6.24
	30	5.34	4.88	5.18	5.44	4.86	7.28
PROC MIXED TOEP	10	4.94	5.68	5.52	5.66	5.52	5.76
	20	4.86	4.98	4.96	5.64	4.60	5.16
	30	5.36	5.14	5.10	5.30	4.50	5.36
PROC MIXED UN	10	5.20	5.92	4.74	4.94	4.96	4.58
	20	4.90	4.86	5.46	5.56	4.80	5.50
	30	5.32	5.54	5.16	5.26	4.56	5.40
PROC MIXED AR(1)	10	4.36	4.50	4.80	4.32	3.78	4.38
	20	4.70	5.00	5.04	5.58	3.74	4.52
	30	5.40	4.86	4.84	5.30	3.92	4.88

Table B7. Type I Error for Treatment Effect when Effect Size on Time is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	4.56	4.70	4.84	4.88	5.10	4.84
	20	4.32	4.60	5.32	5.48	5.10	5.30
	30	5.06	5.10	5.30	4.94	5.02	5.00
PROC MIXED CS	10	4.56	4.70	4.84	4.88	5.10	4.84
	20	4.32	4.60	5.32	5.48	5.10	5.30
	30	5.06	5.10	5.30	4.94	5.02	5.00
PROC MIXED TOEP	10	4.74	5.12	4.90	5.24	5.18	4.96
	20	4.40	4.48	5.36	5.32	5.10	5.48
	30	5.08	5.24	5.32	4.92	4.98	5.14
PROC MIXED UN	10	4.56	4.72	4.84	4.88	5.10	4.84
	20	4.32	4.60	5.32	5.48	5.10	5.30
	30	5.06	5.10	5.30	4.94	5.02	5.00
PROC MIXED AR(1)	10	5.14	4.76	5.00	5.46	6.32	6.82
	20	4.28	4.68	5.62	5.60	6.00	7.14
	30	5.10	4.82	5.08	5.14	5.68	6.78

Table B8. Type I Error for Interaction Effect when Effect Size on Time is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.38	5.42	5.66	5.26	4.74	4.72
	20	4.90	4.92	5.34	5.04	4.90	5.24
	30	4.44	5.08	4.98	5.28	4.78	4.72
PROC MIXED CS	10	5.66	5.04	6.00	5.44	5.34	6.60
	20	4.70	4.86	4.92	4.84	5.14	6.62
	30	4.52	5.16	5.20	5.76	5.42	5.80
PROC MIXED TOEP	10	6.10	6.36	6.16	5.92	5.12	5.70
	20	4.82	4.82	4.46	4.52	4.84	4.98
	30	4.22	5.20	5.04	5.14	4.90	4.98
PROC MIXED UN	10	5.38	5.42	5.66	5.26	4.74	4.72
	20	4.90	4.92	5.34	5.04	4.90	5.24
	30	4.44	5.08	4.98	5.28	4.78	4.72
PROC MIXED AR(1)	10	5.48	5.10	5.22	4.94	3.86	3.54
	20	4.62	4.68	4.48	4.66	4.10	4.14
	30	4.30	5.18	5.02	5.54	4.10	4.52

Table B9. Type I Error for Treatment Effect when Effect Size on Time is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	4.54	5.16	5.28	4.72	4.86	5.18
	20	5.12	4.68	4.82	5.04	5.54	5.06
	30	5.00	5.02	4.46	4.82	5.16	4.62
PROC MIXED CS	10	4.54	5.16	5.28	4.72	4.86	5.18
	20	5.12	4.68	4.82	5.04	5.54	5.06
	30	5.00	5.02	4.46	4.82	5.16	4.62
PROC MIXED TOEP	10	4.54	5.32	5.50	4.96	4.98	5.16
	20	5.28	4.56	4.94	4.88	5.68	5.04
	30	4.88	4.86	4.48	4.96	5.12	4.66
PROC MIXED UN	10	4.54	5.16	5.28	4.72	4.86	5.18
	20	5.12	4.68	4.82	5.04	5.54	5.06
	30	5.00	5.02	4.46	4.82	5.16	4.62
PROC MIXED AR(1)	10	4.46	4.68	5.28	5.54	6.06	7.30
	20	5.12	4.60	5.26	5.14	6.74	6.70
	30	4.86	4.94	4.68	5.12	6.10	6.26

Table B10. Type I Error for Interaction Effect when Effect Size on Time is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.76	4.72	4.72	4.58	4.94	4.44
	20	5.16	4.68	5.06	5.08	5.28	4.76
	30	4.76	5.22	4.62	4.76	4.96	4.62
PROC MIXED CS	10	4.52	5.40	4.92	5.16	5.78	6.88
	20	5.44	5.22	4.70	5.50	5.54	6.58
	30	5.02	5.18	4.70	5.16	5.46	6.10
PROC MIXED TOEP	10	5.16	5.94	5.22	5.02	5.42	5.80
	20	5.38	5.16	4.74	5.06	5.50	4.88
	30	5.08	5.80	4.92	4.78	5.14	4.44
PROC MIXED UN	10	4.76	4.72	4.72	4.58	4.94	4.44
	20	5.16	4.68	5.06	5.08	5.28	4.76
	30	4.76	5.22	4.62	4.76	4.96	4.62
PROC MIXED AR(1)	10	4.22	5.02	4.64	4.82	3.66	4.12
	20	5.48	5.00	4.26	5.20	4.42	4.10
	30	4.90	5.40	4.62	4.50	4.50	4.10

Table B11. Type I Error for Treatment Effect when Effect Size on Time is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	5.16	4.92	5.04	4.68	4.82	5.36
	20	4.74	5.12	4.82	5.20	4.64	4.62
	30	5.22	4.80	4.90	4.76	5.10	4.78
PROC MIXED CS	10	5.16	4.92	5.04	4.68	4.82	5.36
	20	4.74	5.12	4.82	5.20	4.64	4.62
	30	5.22	4.80	4.90	4.76	5.10	4.78
PROC MIXED TOEP	10	5.38	5.10	5.02	4.88	4.92	5.32
	20	4.64	5.28	4.88	5.08	4.70	4.62
	30	5.20	4.88	4.92	4.74	4.98	4.74
PROC MIXED UN	10	5.16	4.92	5.04	4.68	4.82	5.36
	20	4.74	5.12	4.82	5.20	4.64	4.62
	30	5.22	4.80	4.90	4.76	5.10	4.78
PROC MIXED AR(1)	10	5.06	4.92	5.30	5.16	6.00	7.20
	20	4.60	5.46	4.92	5.06	5.88	6.40
	30	5.22	4.80	5.12	5.00	5.86	6.08

Table B12. Type I Error for Interaction Effect when Effect Size on Time is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.16	4.76	4.56	5.80	5.14	5.14
	20	5.32	4.80	5.50	5.28	4.30	5.00
	30	4.90	5.38	5.14	5.20	4.60	5.38
PROC MIXED CS	10	5.18	5.24	4.68	5.98	5.86	6.60
	20	5.14	5.36	5.40	5.32	4.94	6.60
	30	4.86	5.38	5.60	4.78	5.68	7.54
PROC MIXED TOEP	10	5.18	6.12	5.16	6.40	5.02	5.88
	20	5.26	5.10	5.36	4.98	4.60	4.76
	30	5.10	5.34	5.42	4.64	4.56	5.58
PROC MIXED UN	10	5.16	4.76	4.56	5.82	5.14	5.14
	20	5.32	4.80	5.50	5.28	4.30	5.00
	30	4.90	5.38	5.14	5.20	4.60	5.38
PROC MIXED AR(1)	10	5.06	5.06	4.62	5.60	3.46	3.72
	20	5.30	5.22	5.04	4.74	3.94	3.92
	30	4.88	5.38	5.30	4.66	4.36	5.16

Table B13. Type I Error for Treatment Effect when Effect Size on Interaction is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	5.06	5.14	4.48	4.76	4.80	5.06
	20	4.84	5.20	4.82	5.06	4.38	5.30
	30	4.86	4.96	5.10	5.12	4.90	4.90
PROC MIXED CS	10	5.06	5.14	4.48	4.76	4.80	5.06
	20	4.84	5.20	4.82	5.06	4.38	5.30
	30	4.86	4.96	5.10	5.12	4.90	4.90
PROC MIXED TOEP	10	5.22	5.46	4.40	4.78	5.08	4.92
	20	5.00	5.00	4.72	5.10	4.50	5.40
	30	4.84	5.00	5.08	4.94	4.98	5.02
PROC MIXED UN	10	5.06	5.14	4.48	4.76	4.80	5.06
	20	4.84	5.20	4.82	5.06	4.38	5.30
	30	4.86	4.96	5.10	5.12	4.90	4.90
PROC MIXED AR(1)	10	5.04	5.10	4.88	4.98	6.16	7.12
	20	4.80	4.96	5.24	5.32	5.64	7.44
	30	4.92	5.30	5.50	5.32	5.86	6.78

Table B14. Type I Error for Time Effect when Effect Size on Interaction is $(\frac{1}{2})\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.98	5.02	5.16	5.58	4.74	5.02
	20	4.84	5.04	4.96	4.46	4.56	4.92
	30	5.00	5.20	4.96	5.70	5.28	4.58
PROC MIXED CS	10	4.90	4.88	5.16	5.24	5.50	6.78
	20	4.80	4.78	5.36	5.22	4.82	6.96
	30	4.74	4.98	4.86	5.44	6.38	6.28
PROC MIXED TOEP	10	4.94	5.82	5.30	5.50	4.82	6.00
	20	4.70	4.72	4.96	4.88	4.88	5.22
	30	4.70	5.16	4.64	5.36	5.40	5.22
PROC MIXED UN	10	4.98	5.02	5.16	5.58	4.74	5.02
	20	4.48	5.04	4.96	4.46	4.56	4.92
	30	5.00	5.20	4.96	5.70	5.28	4.58
PROC MIXED AR(1)	10	4.74	4.66	4.52	4.46	3.44	4.06
	20	4.62	4.52	5.10	4.46	4.02	3.98
	30	4.70	5.06	4.70	4.88	4.68	4.32

Table B15. Type I Error for Treatment Effect when Effect Size on Interaction is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM UV F test	10	5.06	4.58	4.66	4.94	5.04	5.20
	20	4.80	4.94	4.80	5.18	4.82	5.02
	30	5.28	4.80	4.86	4.98	4.70	5.22
PROC MIXED CS	10	5.06	4.58	4.66	4.94	5.04	5.20
	20	4.80	4.94	4.80	5.18	4.82	5.02
	30	5.28	4.80	4.86	4.98	4.70	5.22
PROC MIXED TOEP	10	5.28	4.86	4.70	4.92	5.14	5.08
	20	4.94	4.90	4.74	5.32	4.82	5.22
	30	5.36	5.00	4.92	4.90	4.76	5.14
PROC MIXED UN	10	5.06	4.58	4.66	4.94	5.04	5.20
	20	4.80	4.94	4.80	5.18	4.82	5.02
	30	5.28	4.80	4.86	4.98	4.70	5.22
PROC MIXED AR(1)	10	5.28	4.74	5.04	5.04	6.32	6.70
	20	5.04	5.10	4.90	5.20	5.62	6.90
	30	5.52	4.90	5.04	5.68	5.54	6.92

Table B16. Type I Error for Time Effect when Effect Size on Interaction is σ_s

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.40	4.46	4.96	4.66	5.04	5.26
	20	4.74	4.92	5.34	5.30	4.56	5.40
	30	4.62	5.08	4.78	4.94	5.26	4.74
PROC MIXED CS	10	5.28	5.12	4.50	5.06	5.86	7.04
	20	4.72	5.10	5.44	5.46	5.48	7.04
	30	4.82	5.10	5.54	4.78	5.84	6.50
PROC MIXED TOEP	10	5.58	5.58	5.46	5.68	5.20	5.98
	20	4.92	5.16	5.34	5.32	4.84	6.00
	30	4.70	5.02	5.32	4.74	5.16	4.84
PROC MIXED UN	10	5.40	4.46	4.96	4.66	5.04	5.26
	20	4.74	4.92	5.34	5.30	4.56	5.40
	30	4.62	5.08	4.78	4.94	5.26	4.74
PROC MIXED AR(1)	10	5.02	5.08	4.34	4.72	3.76	4.14
	20	4.88	5.16	5.08	5.20	4.12	4.78
	30	4.86	4.82	5.32	4.54	4.56	4.08

Table B17. Type I Error for Treatment Effect when Effect Size on Interaction is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	4.18	5.06	4.60	5.50	5.34	5.06
	20	4.78	5.06	4.92	5.40	4.68	4.70
	30	5.34	5.06	5.16	4.60	4.76	5.10
PROC MIXED CS	10	4.18	5.06	4.60	5.50	5.34	5.06
	20	4.78	5.06	4.92	5.40	4.68	4.70
	30	5.34	5.06	5.16	4.60	4.76	5.10
PROC MIXED TOEP	10	4.20	5.30	4.70	5.42	5.24	5.00
	20	4.72	5.04	4.92	5.32	4.82	4.80
	30	5.18	5.04	5.22	4.64	4.66	5.08
PROC MIXED UN	10	4.18	5.04	4.60	5.50	5.34	5.06
	20	4.78	5.06	4.92	5.40	4.68	4.70
	30	5.34	5.06	5.16	4.60	4.76	5.10
PROC MIXED AR(1)	10	4.16	4.82	4.72	5.48	6.60	7.04
	20	4.76	4.94	5.28	5.54	5.76	6.46
	30	5.52	5.00	5.44	4.80	5.78	6.74

Table B18. Type I Error for Time Effect when Effect Size on Interaction is $2\sigma_s$

Number of Samples = 5,000							
Method of Analysis	Sample Size	rho = 0		rho = 0.25		rho = 0.75	
		k = 4	k = 6	k = 4	k = 6	k = 4	k = 6
PROC GLM Wilk's Lambda	10	5.04	4.88	4.88	4.86	4.70	5.08
	20	4.72	5.04	5.16	5.04	5.12	5.62
	30	4.92	5.00	5.32	5.34	4.84	5.36
PROC MIXED CS	10	5.40	4.46	5.56	5.12	5.38	6.22
	20	5.12	4.90	5.22	5.80	5.42	6.66
	30	5.06	5.02	5.30	4.94	5.60	6.86
PROC MIXED TOEP	10	5.30	5.32	5.54	5.64	4.98	6.06
	20	4.98	4.70	5.12	5.52	5.14	5.74
	30	5.16	5.00	4.96	4.98	4.96	5.04
PROC MIXED UN	10	5.04	4.88	4.88	4.86	4.70	5.08
	20	4.72	5.04	5.16	5.04	5.12	5.62
	30	4.92	5.00	5.32	5.34	4.84	5.36
PROC MIXED AR(1)	10	4.90	4.42	4.50	4.68	3.80	3.54
	20	4.82	4.94	4.72	5.24	4.12	4.28
	30	5.04	4.88	5.12	4.88	4.20	4.48

Tables B19, B20, and B21 are shown for reference of using fit statistics, or model fitting criteria, to determine which covariance structure to use when testing repeated measures data. The data on these tables is shown as the percentage of times the fit statistic was the lowest for that covariance structure (because smaller means a better fit). For example, on Table B19 when $k = 4$, and $n = 10$, the AIC fit statistics choose AR(1) to be the best fit for the data 40% of the time. Table B19 shows that when correlation is equal to 0, the fit statistics are generally showing that CS has the best fit. This is logical because with a correlation of 0 there is no change in the covariance as time increases, which fits the trend assumed in CS. As shown on Tables B20 and B21, when correlation was greater than zero, AIC, AICC and BIC all indicated that the data had an AR(1) covariance structure. This is good because AR(1) is the true covariance structure of the data. Note that the percentage of time the fit statistics chose AR(1) gets larger as correlation, sample size, and the number of repeated measures increase. Therefore, when researchers have a data set with a large sample size, the fit statistics may be more reliable. The AIC does not penalize complex covariance structures as much as AICC and BIC. Guerin and Stroup (2000) recommend using the AIC fit statistic.

Table B19. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0

Number of Samples = 5,000 Rho = 0							
Tested Covariance Structure	Sample Size	k = 4			k = 6		
		AIC	AICC	BIC	AIC	AICC	BIC
CS	10	38.82	45.38	42.24	38.84	47.76	44.16
TOEP		10.40	7.70	8.58	5.58	6.30	6.50
UN		10.78	0.94	6.26	14.16	0.02	6.78
AR(1)		40.00	45.98	42.92	38.84	45.92	42.56
CS	20	45.20	47.38	50.06	45.66	47.46	50.22
TOEP		8.26	6.82	3.18	8.12	6.72	1.92
UN		4.92	1.86	0.38	2.14	0.14	0
AR(1)		41.62	43.94	46.38	44.08	45.68	47.86
CS	30	44.24	46.06	49.88	47.16	48.20	50.70
TOEP		8.94	7.90	2.04	7.44	6.08	0.80
UN		4.46	2.14	0.10	0.98	0.16	0
AR(1)		42.36	47.98	47.98	44.42	45.56	48.50

Table B20. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0.25

Number of Samples = 5,000 Rho = 0.25							
Tested Covariance Structure	Sample Size	k = 4			k = 6		
		AIC	AICC	BIC	AIC	AICC	BIC
CS	10	33.80	39.46	36.62	27.52	33.42	31.12
TOEP		10.44	8.22	9.16	10.12	7.50	7.94
UN		11.70	0.76	6.74	15.22	0.14	6.80
AR(1)		44.06	51.56	47.48	47.14	58.94	54.14
CS	20	31.10	32.70	34.60	23.44	24.40	25.66
TOEP		9.52	8.52	3.86	9.28	7.62	2.28
UN		4.68	1.42	0.32	2.16	0.06	0
AR(1)		54.70	57.36	61.22	65.12	67.92	72.06
CS	30	26.46	27.14	29.16	16.96	17.36	18.42
TOEP		9.06	8.00	2.24	8.42	7.32	0.74
UN		3.80	1.86	0.04	0.90	0.14	0
AR(1)		60.68	63.00	68.56	73.72	75.18	80.84

Table B21. Fit Statistics for data with an Effect Size of $(\frac{1}{2})\sigma_s$ on Treatment – Rho = 0.75

Number of Samples = 5,000 Rho = 0.75							
Tested Covariance Structure	Sample Size	k = 4			k = 6		
		AIC	AICC	BIC	AIC	AICC	BIC
CS	10	29.04	33.32	31.22	17.48	21.84	19.84
TOEP		8.94	6.62	7.46	9.72	7.12	7.50
UN		10.44	0.84	6.00	15.26	0.08	7.36
AR(1)		51.58	59.22	55.32	57.54	70.96	65.30
CS	20	24.70	26.16	27.72	10.34	10.88	12.44
TOEP		8.60	7.24	3.02	15.04	12.40	3.62
UN		4.84	1.62	0.24	2.38	0.16	0.04
AR(1)		61.86	64.98	69.02	72.24	76.56	83.90
CS	30	21.18	22.16	24.94	4.36	4.62	7.82
TOEP		11.28	10.04	2.60	22.76	20.66	3.22
UN		4.62	2.00	0.14	1.80	0.28	0
AR(1)		62.92	65.80	72.32	71.08	74.44	88.96

APPENDIX C. CODE USED FOR EXAMPLE

Below is the code used to analyze the example shown in Chapter 5 of this paper. As mentioned in Chapter 5, this code and methodology follows from Littell et al. (2006a).

```
options ls=80 ps=120;

title1 'Mice Example';

data mice_uv;          *** Univariate data structure ***;
  infile 'mice_uvdataRB.txt' dlm='09'x firstobs=2;
  input subj_id trt$ week weight @@;
  ;;;;

proc print data=mice_uv;    *** checking the data ***;
  title2 'UV Mice Data';
  run;

proc mixed data=mice_uv;
  class trt week subj_id;
  model weight = trt|week/ddfm=kr;
  repeated week/subject=subj_id(trt) type=UN r rcorr;
  run;

proc mixed data=mice_uv;
  class trt week subj_id;
  model weight = trt|week/ddfm=kr;
  repeated week/subject=subj_id(trt) type=UN sscp rcorr;
  ods output covparms = cov rcorr = corr;
  run;

data weeks;          *** Calculating lag for UN var and cov estimates ***;
  do week1=1 to 8;
    do week2=1 to week1;
      lag=week1-week2;
      output;
    end;
  end;
  run;

data covplot;
  merge weeks cov;
  run;

proc print data=covplot;
  run;

axis1 order = (2 to 12.99 by 1.57)
  label = (angle=90 h=2 'Covariance of Between Subject Effects');
axis2 order = (0 to 7 by 1)
  minor = none
  label = (h=2 'Lag in Weeks');
legend value=(h=2)
  label = ('From Week')
  mode = protect
```

```

        position = (top right inside);
symbol11 color=black interpol=join line =1 value =circle;
symbol12 color=black interpol=join line =2 value =dot;
symbol13 color=black interpol=join line =5 value =triangle;
symbol14 color=black interpol=join line =3 value =square;
symbol15 color=black interpol=join line =8 value =plus;
symbol16 color=black interpol=join line =10 value =asterisk;
symbol17 color=black interpol=join line =23 value =star;
symbol18 color=black interpol=join line =27 value =cross;

proc gplot data=covplot; *** Plotting lag of UN cov structure ***;
    plot estimate*lag=week2/ vaxis=axis1 haxis=axis2 legend=legend1;
    title2 'Plot of Covariance of Between-Subjects vs Lag in Weeks';
    run;
*****
*****Plotting Covariance strucutres to compare to UN plot*****
*****and obtaining Fit Statistic values and Test P-values*****
*****;
ods listing close;
ods output
    r = UNr
    CovParms=CovParmsUN
    Tests3=Tests3UN
    FitStatistics=FitUN;
proc mixed data=mice_uv ic; *** obtaing the UN data ***;
    class trt week subj_id;
    model weight = trt|week/ddfm=kr;
    repeated week/subject=subj_id(trt) type=UN r;
    run;
ods listing;
data UNdata;
    set UNr;
    week1=row;
    array col (8) col1-col8;
    do i=1 to 8;
        week2=i;
        lag=week1-week2;
        if lag>=0;
            cov= col(i);
            structure="UN";
            output;
        end;
    keep week1 week2 lag cov structure;
    run;
/*proc print data=UNdata;*/
/* title2 'UN Covariance';*/
/* run;*/

ods listing close;
ods output
    r = CSr
    CovParms=CovParmsCS
    Tests3=Tests3CS
    FitStatistics=FitCS;
proc mixed data=mice_uv ic; *** obtaing the CS data ***;
    class trt week subj_id;

```

```

        model weight = trt|week/ddfm=kr;
        repeated week/subject=subj_id(trt) type=CS r;
    run;
ods listing;
data CSdata;
    set CSr;
    week1=row;
    array col (8) col1-col8;
    do i=1 to 8;
        week2=i;
        lag=week1-week2;
        if lag>=0;
        cov= col(i);
        structure="CS";
        output;
    end;
    keep week1 week2 lag cov structure;
    run;
/*proc print data=CSdata;*/
/*    title2 'CS Covariance';*/
/*    run;*/

ods listing close;
ods output
    r = ARr
    CovParms=CovParmsAR
    Tests3=Tests3AR
    FitStatistics=FitAR;
proc mixed data=mice_uv ic;    *** obtaing the AR(1) data ***;
    class trt week subj_id;
    model weight = trt|week/ddfm=kr;
    repeated week/subject=subj_id(trt) type=AR(1) r;
    run;
ods listing;
data ARdata;
    set ARr;
    week1=row;
    array col (8) col1-col8;
    do i=1 to 8;
        week2=i;
        lag=week1-week2;
        if lag>=0;
        cov= col(i);
        structure="AR";
        output;
    end;
    keep week1 week2 lag cov structure;
    run;
/*proc print data=ARdata;*/
/*    title2 'AR Covariance';*/
/*    run;*/

ods listing close;
ods output
    r = TOEPr
    CovParms=CovParmsTOEP

```



```

Tests3=Tests3TOEP
FitStatistics=FitTOEP;
proc mixed data=mice_uv ic;   *** obtaing the TOEP data ***;
class trt week subj_id;
model weight = trt|week/ddfm=kr;
repeated week/subject=subj_id(trt) type=TOEP r;
run;
ods listing;
data TOEPdata;
set TOEP;
week1=row;
array col (8) col1-col8;
do i=1 to 8;
    week2=i;
    lag=week1-week2;
    if lag>=0;
    cov= col(i);
    structure="TOEP";
    output;
end;
keep week1 week2 lag cov structure;
run;
/*proc print data=TOEPdata;*/
/* title2 'TOEP Covariance';*/
/* run;*/

data CovLagData;
set UNdata CSdata ARdata TOEPdata;
run;

proc sort data=CovLagData;
by structure lag;
run;

/*proc print data=CovLagData;*/
/* title2 'Covariance Data by Covariance Structures';*/
/* run;*/

symbol1 color=black interpol=join value =circle;
symbol2 color=black interpol=join value =dot;
symbol3 color=black interpol=join value =triangle;
symbol4 color=black interpol=join value =square;
symbol5 color=black interpol=join line =1 value =plus;
symbol6 color=black interpol=join line =2 value =asterisk;
symbol8 color=black interpol=join line =3 value =cross;

proc gplot data=CovLagData;   *** Plotting the Cov vs Lag for each cov.
Structure ***;
by structure;
plot cov*lag=week2;
title2 'Plot of Covariance Structures vs Lag in Weeks';
run;

data Tests3UN;
set Tests3UN;
structure="UN";

```

```

run;

data Tests3CS;
  set Tests3CS;
  structure="CS";
run;

data Tests3AR;
  set Tests3AR;
  structure="AR";
run;

data Tests3TOEP;
  set Tests3TOEP;
  structure="TOEP";
run;

data Tests3Data;  *** combining the p-value data ***;
  set Tests3TOEP Tests3UN Tests3CS Tests3AR ;
run;

proc sort data=Tests3Data;
  by Effect structure;
run;

proc print data=Tests3Data; *** printing the p-values by cov. Structure ***;
  title2 'P-values and F-values for Type 3 Tests by Covariance
Structure';
run;

data FitUN;
  set FitUN;
  structure="UN";
run;

data FitCS;
  set FitCS;
  structure="CS";
run;

data FitAR;
  set FitAR;
  structure="AR";
run;

data FitTOEP;
  set FitTOEP;
  structure="TOEP";
run;

data FitData;  *** combining the fit statistic data ***;
  set FitTOEP FitUN FitCS FitAR ;
run;

proc sort data=FitData;
  by Descr;

```

```
run;  
  
proc print data=FitData; *** printing the fit statistics data ***;  
  title2 'Fit Statistics by Covariance Structure';  
run;
```

APPENDIX D. FURTHER POWER ANALYSIS AS CORRELATION INCREASES

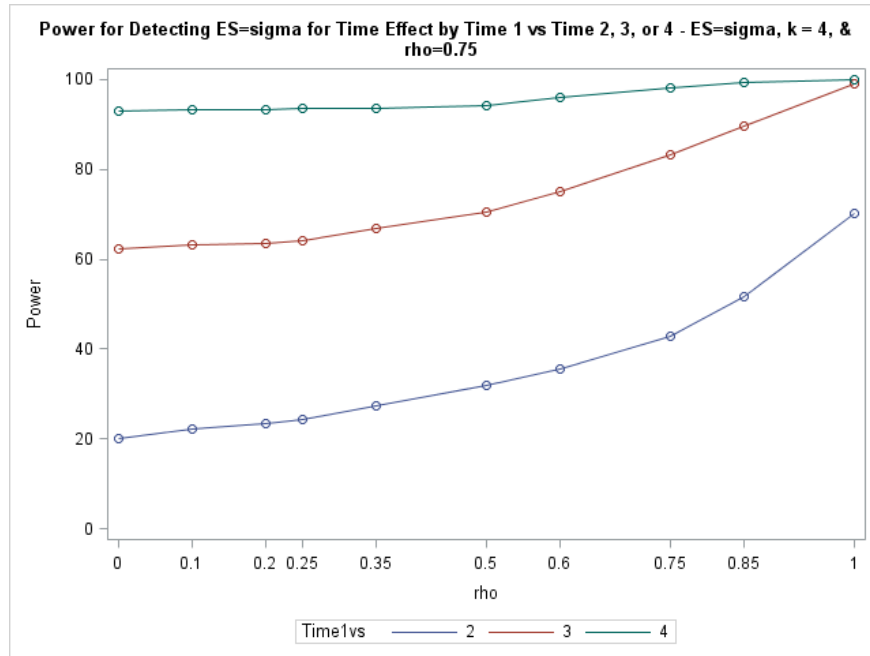


Figure D1. Power for Time 1 vs Time 2, 3, and 4 – k = 4

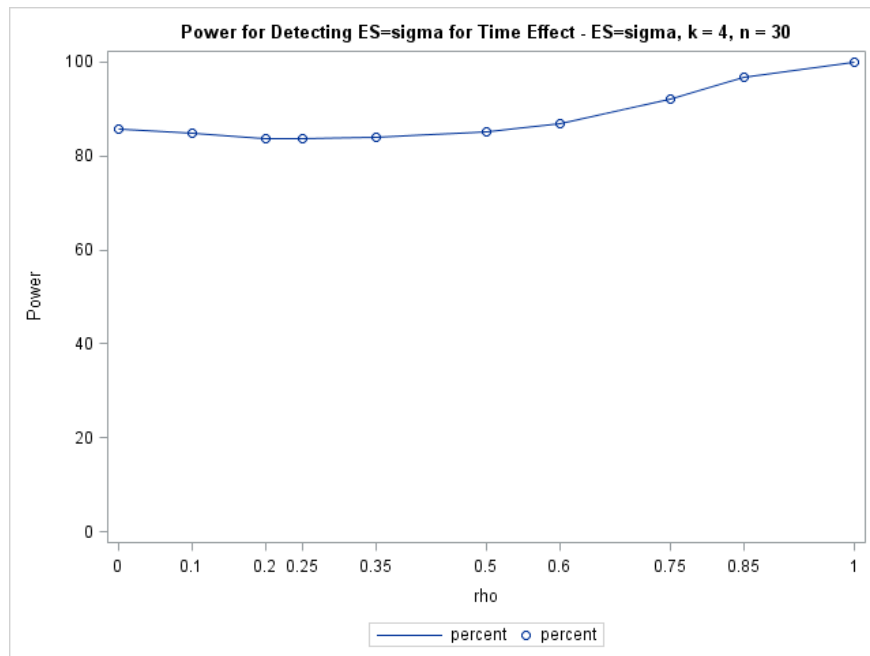


Figure D2. Power for Time Effect – k = 4

Figure D1 shows the pairwise comparisons of power for Time 1 vs. Time 2, Time 1 vs. Time 3, and Time 1 vs. Time 4. Here the power increases for all comparisons as correlation increases, but increases faster for times further away from time 1 because the effect size is larger.

Figure D2 shows the overall power plot as for the time effect as correlation increases when $k = 4$, $n = 30$, effect size = σ_s , and AR(1) is the covariance structure specified in PROC MIXED. The power decreases slightly from correlation of 0 to 0.25, and then increases after correlation equals 0.25. This unexpected decrease may have something to do with the small sample size and the way the effect size increments in beta when simulating the data.

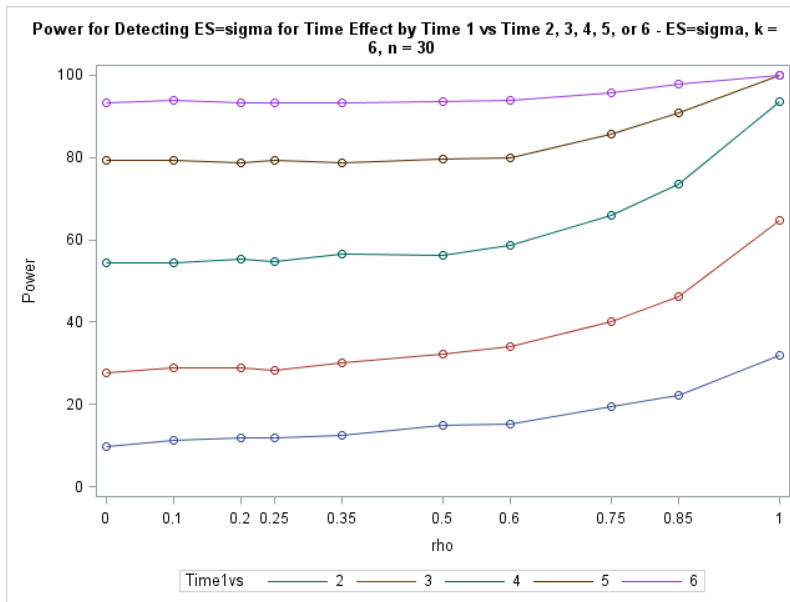


Figure D3. Power for Time 1 vs Time 2, 3, 4, 5, and 6 – $k = 6$

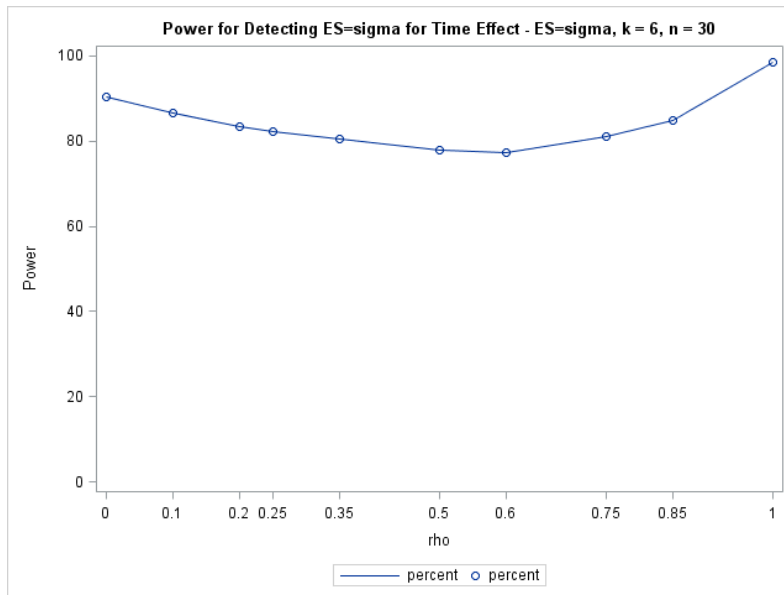


Figure D4. Power for Time Effect – $k = 6$

Figure D3 shows the pairwise comparisons of power for Time 1 vs. Time 2, Time 1 vs. Time 3, Time 1 vs. Time 4, Time 1 vs. Time 5, and Time 1 vs. Time 6. Here the power also increases for all comparisons as correlation increases, and increases faster for times further away from time 1 because the effect size is larger. Figure D4 shows the overall power plot as for time effect as correlation increases when $k = 6$, $n = 30$, effect size = σ_s , and AR(1) is the covariance structure specified in PROC MIXED. The power decrease is more obvious in Figure D4 than in Figure D2. In Figure D4 the power decreases from correlation of 0 to 0.60, and then increases after correlation equals 0.60. Just as stated before, this unexpected decrease may have something to do with the small sample size. Also, because the effect size increments with smaller units when $k = 6$, this could be why the decrease is more drastic than the decrease seen in power when $k = 4$.

Note that different code was used to obtain the p-values for the pairwise estimates shown in Figure D1, D2, D3, and D4. See Figure D5 below to reference the code that was used to obtain the pairwise p-values when $k=4$.

```

proc mixed data=rm.uv_AR1_4_10;
  by sample;
  class trt period subj_id;
  model stress = trt | period / ddfm=kr;
  repeated period / subject=subj_id type=AR(1);
  estimate 'time 1 vs 2 for Trt 1'
    period 1 -1 0 0;
  estimate 'time 1 vs 3 for Trt 1'
    period 1 0 -1 0;
  estimate 'time 1 vs 4 for Trt 1'
    period 1 0 0 -1;
  title2 "Repeated Measures ANOVA using Mixed Model
  Approach -- AR(1)";
run;

```

Figure D5. Code to Obtain P-values for Pairwise Comparisons