A Three-Paper Dissertation on Longitudinal Data Analysis in Education and Psychology

Hedyeh Ahmadi

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

A Three-Paper Dissertation on Longitudinal Data Analysis in Education and Psychology

Hedyeh Ahmadi

In longitudinal settings, modeling the covariance structure of repeated measure data is essential for proper analysis. The first paper in this three-paper dissertation presents a survey of four journals in the fields of Education and Psychology to identify the most commonly used methods for analyzing longitudinal data. It provides literature reviews and statistical details for each identified method. This paper also offers a summary table giving the benefits and drawbacks of all the surveyed methods in order to help researchers choose the optimal model according to the structure of their data. Finally, this paper highlights that even when scholars do use more advanced methods for analyzing repeated measure data, they very rarely report (or explore in their discussions) the covariance structure implemented in their choice of modeling. This suggests that, at least in some cases, researchers may not be taking advantage of the optimal covariance patterns. This paper identifies a gap in the standard statistical practices of the fields of Education and Psychology, namely that researchers are not modeling the covariance structure as an extension of fixed/random effects modeling. The second paper introduces the General Serial Covariance (GSC) approach, an extension of the Linear Mixed Modeling (LMM) or Hierarchical Linear Model (HLM) techniques that models the covariance structure using spatial correlation functions such as Gaussian, Exponential, and other patterns. These spatial correlations model the covariance structure in a continuous manner and therefore can deal with missingness and imbalanced data in a straightforward way. A simulation study in the second paper reveals that when data are consistent with

the GSC model, using basic HLMs is not optimal for the estimation and testing of the fixed effects. The third paper is a tutorial that uses a real-world data set from a drug abuse prevention intervention to demonstrate the use of the GSC and basic HLM models in R programming language. This paper utilizes variograms (a visualization tool borrowed from geostatistics) among other exploratory tools to determine the covariance structure of the repeated measure data. This paper aims to introduce the GSC model and variogram plots to Education and Psychology, where, according to the survey in the first paper, they are not in use. This paper can also help scholars seeking guidance for interpreting the fixed effect-parameters.

Table of Contents

ACKNOWLEDGMENTS

I feel tremendously fortunate to have the unflagging support of my academic advisors, my family, and my friends who always believed in me and encouraged me to pursue my academic ambitions. I want to gratefully acknowledge all of my academic and non-academic mentors (whose names may not be listed here), who throughout the years have guided me along the way. I will forever be their student.

I would like to first and foremost thank my wonderful husband who has been there for me in the past 12 years, supporting me in every step of my goals, academic and otherwise. He is the wisest and best cheerleader I could ask for.

I would like to also thank my amazing advisor, Prof. Elizabeth Tipton, who despite her very busy schedule has been there for me at every step—from situating me in the department, encouraging me to take on jobs I thought were beyond my abilities, stopping me when I was going overboard, and sending me the "hang in there" emails that meant so much to me. She is the mentor and role model that every graduate student dreams of.

I would like to also acknowledge Prof. Bryan Keller for being so generous with his time and providing the data set for this dissertation. His sage advice constantly shed light on this project. In addition, I would like to thank Prof. Matthew Johnson for his valuable feedback that always took my dissertation to the next level. I would like to also express my gratitude to the rest of my committee, Prof. Oren Pizmony-Levy and Prof. Minjeong Jeon, who both graciously provided helpful comments. In addition, I would like to extend my sincere thanks to Prof. Daniel Gillen for his informative lecture notes that provided the seed for this dissertation.

I feel enormously lucky to have my mom and my in-laws, who patiently waited all these years for this moment, sharing all the happy and the more trying days and providing invaluable emotional and professional support. I would like to thank the rest of my family, who always lent me an ear when I needed love and support. Finally, I'm grateful for the memory and lessons of my dad, who I know would be proud today.

To my friends (you know who you are!), who listened uncomplainingly in my venting moments, who understood when I missed a lot of their special occasions to stay on track with my goals, and who always reminded me that "this shall pass"—thank you.

Last but not least, I would be remiss if I did not mention the fabulous Columbia University campus, with all of its beautiful libraries and cafeterias that sheltered me on all those cold and snowy days and helped me to find a new place to study every month!

PREFACE

This dissertation consists of three papers that are intended to be submitted to three different journals. It therefore departs from the standard dissertation structure of sequential chapters. Each paper has its own abstract, table of contents, main body, discussion, appendix, and references in accordance with Columbia University's formatting guidelines. The structure of the three papers may change upon submission to the respective relevant journals, so readers interested in the most updated version of the three papers can contact me for updated versions of the papers.

Paper 1: A Comprehensive Review of Methods for Analyzing Repeated Measure Data in Education and Psychology

Hedyeh Ahmadi

Teachers College, Columbia University

ABSTRACT

Paper 1: A Comprehensive Review of Methods for Analyzing Repeated Measure Data in Education and Psychology

Hedyeh Ahmadi

This paper presents a comprehensive review of longitudinal data analysis methods in both qualitative and quantitative formats. The qualitative review can help researchers find examples of methods of interest in the Education and Psychology literature. The quantitative survey and methods sections can help researchers to identify the most commonly used methods in specific journals and in these disciplines overall. For each journal, detailed statistical summaries, including frequency of each method, sample sizes, and number of repeated measurements per study, are provided as well. Recommendations are offered based on these observations that can improve the data collection and analysis methods in Education and Psychology research. The longitudinal methods are surveyed and broken down into two categories to demonstrate how many researchers continue to use *traditional* methods with rigid and unrealistic assumptions when *advanced* models can offer improved statistical properties and more realistic assumptions. To better understand the strengths and limitations of each method, this paper also presents a brief statistical methods section for every model reviewed. A summary table of all the reviewed methods is also presented to help scholars consider the pros and cons of each approach and select the optimal method.

Table of Contents

A Comprehensive Review of Methods for Analyzing Repeated Measure Data

1. Introduction to the Review of Longitudinal Data Analysis in Education and Psychology

Longitudinal analyses are studies in which the response of the same individual is measured on multiple occasions (Fitzmaurice, Laird, & Ware, 2004). Therefore, the independent assumption of observations in longitudinal studies is violated. In modeling these types of data, the researcher needs to account for potential correlation within each subject's measurement and between-subject heterogeneity. Assessing longitudinal data allows researchers to:

1. Investigate changes of outcome(s) over time (i.e., whether/how individuals change over time) and their relations to study variables of interest,
2. Examine interindividual similarities/differences (i.e., whether individuals' respective changes are similar or different),
3. Make claims about causal effects with a better statistical foundation than cross-sectional studies allow for (Fitzmaurice et al., 2004; Gustafsson, 2010; Liang & Zeger, 1993).

Although longitudinal studies offer more information than cross-sectional studies, there are challenges inherent in longitudinal data. These include:

1. Population heterogeneity that leads to subject-specific deviations from the overall trend in response,
2. Correlated errors of measurement due to close measurement intervals,
3. Presence of missing data due to subjects not remaining for the entire study,
4. Irregularly spaced measurement occasions due to dropout or different individuals' availability,

5. An additional source of correlation caused by the clustering of individuals, such as in schools and classrooms (Gibbons, Hedeker, & DuToit, 2010; Verbeke, Fieuws, Molenberghs, & Davidian, 2014).

To overcome the challenges of longitudinal data and accommodate the complications that may arise during data analysis, a variety of models has been introduced in the statistical literature during the last few decades (Verbeke et al., 2014).

Over time, many different methods and models have been developed to address these various problems. However, methods addressing one problem may not address another problem, and methods developed in one field may predominate there while rarely being applied elsewhere. For this reason, this paper provides a review of methodological developments and models related to longitudinal models. To begin, the longitudinal research literature was reviewed in order to identify the commonly used methods for analyzing longitudinal data quantitatively. This literature included Liang and Zeger (1993); Muthén and Curran (1997); Verbeke and Molenberghs (2009); Diggle, Heagerty, Liang, and Zeger (2002); Menard (2002); Twisk (2003); Fitzmaurice, Laird, and Ware (2004); Molenberghs and Verbeke (2005); Hedeker and Gibbons, (2006); Gibbons, Hedeker, and DuToit (2010); Gustafsson (2010); and Verbeke, Fieuws, Molenberghs, and Davidian (2014).

According to the results of the review, several analytical approaches were identified and categorized into two broad classes, namely, *traditional* versus *advanced*. Traditional techniques include paired t-test, analysis of covariance (ANCOVA), analysis of variance (ANOVA), regression analysis, derived variable approach, and repeated measures univariate/multivariate analysis of variance (RM ANOVA, MANOVA). More advanced analytical techniques include mixed effects modeling (including multilevel modeling, heterogeneity models, and generalized linear mixed models), marginal models using generalized estimating equations (GEE), conditional models (specifically, the transition models), autoregressive models and latent growth curve modeling within the structural equation modeling (SEM) framework, mixture models, time series analysis, non-linear, and non-parametric modeling.

This review focuses on the fields of Education and Psychology. The categories developed above were used to classify relevant longitudinal studies in Education and Psychology (i.e., studies in which the dependent variable(s) have been quantitatively measured on the same subject(s) on two occasions or more, irrespective of the length of the study). Electronic searches of the analysis approaches for longitudinal data in Education and Psychology were then conducted via Google Scholar (DeGraff, DeGraff, & Romesburg, 2013; Martin-Martin, Orduna-Malea, Harzing, & Delgado López-Cózar, 2017) using the previous list of statistical methods in combination with terms such as *education*, *psychology*, *repeated measures*, and *longitudinal data* as key words. The search covered the period 2008 to 2018 (i.e., a period of 11 years). Identified studies were reviewed to determine whether they were in fact longitudinal studies and, if so, which methods were employed to analyze the longitudinal data. Studies conducted in disciplines other than Education and Psychology were excluded. Furthermore, because of restrictions in length, only a selection of these publications are examined in detail in the following review sections.

According to this review, while longitudinal models arise frequently in Education and Psychology, and many different models are readily available, the decision to use one model over another is not often explicitly addressed or justified in the literature. The goal of this paper, therefore, is to both provide a review of current practices regarding longitudinal models and to identify the best methods available, highlighting those that are over- and under-used. The paper begins with a survey of longitudinal data analysis in four well-known journals in Education and Psychology. Then, for each statistical method, a general review of longitudinal data in Education and Psychology publications is provided, followed by a section on the statistical details of each reviewed methodology.[1] Finally, after reviewing these available methods, a summary table is provided in the discussion section which is intended to help researchers choose an appropriate model for their data.

2. A Survey of Longitudinal Analysis Related Articles

In order to investigate the gap between current practices (i.e., the most prominent analysis methods for longitudinal data) and all methods available in longitudinal research, a survey was conducted of relevant articles in the following journals:

- *Journal of Research on Educational Effectiveness* (JREE),
- *Journal of Applied Psychology* (JAP),
- *Developmental Psychology* (DP),
- *Educational Evaluation and Policy Analysis* (EEPA).

The survey was conducted through each journal's website. The length of the search (shown in the first column of Table 1) for each journal varied based on library access, number of articles identified, and duration of the journal's existence (for example, JREE has been published since 2008). For JAP, the length of search was set to be 30 years (1988-2018) in order to yield a sufficient number of articles. For JREE and JAP, the search was conducted on December 8, 2018. Due to the large number of articles available in DP and EEPA, the length of the search in each journal was set to the past 10 years (2010-2019). These searches were conducted on March 7, 2019. For all of the journals, the key word used for search was "longitudinal" (in all fields).

Note that for JREE and JAP, all the searched articles were reviewed due to the small number of yielded articles. However, 650 articles for DP and 154 for EEPA fitted the search criteria. These

[1] Note that the main reference(s) for each statistical method can be identified in Table 7 in the Discussion Section. The statistical notations for each method are heavily borrowed from these references.

articles were sorted by date in descending order (the most recent to the least recent) and numbered accordingly (from 1 to 650 for DP and from 1 to 154 for EEPA). Random numbers were generated in EXCEL using the function RAND(). Then the list was sorted ascendingly according to random numbers (smallest to largest). The first 100 articles for each journal were reviewed. The search results are presented in Table 1; note the wide range of terminologies for the same methods in these journals (e.g., multilevel modeling has been called Hierarchical Linear Modeling (HLM), random coefficient, Linear Mixed Modeling (LMM), etc.).

For JREE, 90 articles were reviewed and 13 were determined not to be related to longitudinal analysis. The mixed-effects modeling approach (including HLM (N = 8, 10.4%), mixed-effects models (N = 7, 9.1%), and multilevel modeling (N = 17, 22.1%)) was the most popular analysis method utilized for longitudinal data, accounting for 41.6% (N = 32) of the analysis methods used. Among the 32 articles using mixed-effects modeling approaches, one article, Language and Reading Research Consortium by Arthur and Davis (2016), specified that the within-subjects error covariance matrix was modeled using an independence structure, and the remaining 31 articles did not mention what specific covariance structure(s) were utilized in their data analysis. Note that within the 31 articles that did not specify what specific covariance structure(s) were utilized, three articles, including August, Branum-Martin, Cardenas-Hagan and Francis (2009); Long (2016); and Edmunds et al. (2017), had applied the Huber-White sandwich estimate to obtain cluster-robust standard errors.

For JAP, 82 articles were reviewed and 7 were determined not to be related to longitudinal study. SEM-related approaches were the most popular analysis approach utilized for longitudinal data, accounting for 45.3% of the analysis methods used in JAP. There were 18 articles (24.0%) utilizing mixed-effects modeling approach (including HLM (N = 5, 6.7%), mixed-effects model (N

= 5, 6.7%), multilevel modeling (N = 4, 5.3%), and random-coefficients model (N = 4, 5.3%)). Among these 18 articles utilizing mixed effects modeling approaches, one article by Sitzmann and Ely (2010) had described in detail how the proper error structure of the random effects was identified. In particular, the error structure was compared against the following three covariance structures, including autoregressive and heterogeneous, first-order autoregressive, and unstructured (Sitzmann & Ely, 2010). The change in deviance statistics was used to choose which error pattern leads to an optimal fit. Sitzmann and Ely (2010) had chosen autoregressive and heterogeneous as the covariance structure used in the data analysis. The remaining 17 articles did not mention what specific covariance structure(s) were utilized in their data analysis.

For DP, 100 articles were reviewed and 6 were determined not to be related to longitudinal study. SEM-related methods were the most popular analysis approach utilized for longitudinal data, accounting for 54.3% (N = 51) of the analysis methods used. There were 12 articles utilizing mixed-effects modeling approaches (including HLM (N = 4, 4.3%), mixed-effects models (N = 3, 3.2%), and multilevel modeling (N = 5, 5.3%)). Even though the mixed-effects modeling approach was the second most popular method (tied with Linear Regression), none of the articles specified the covariance structure used in the analysis.

For EEPA, 100 articles were reviewed and 20 were determined not to be related to longitudinal study. Mixed-effects modeling approaches (including HLM (N = 6, 7.5%), mixed-effects models (N = 16, 20.0%), and multilevel modeling (N = 5, 6.3%)) were the most popular analysis method utilized for longitudinal data, accounting for 33.8% (N = 27) of the analysis methods used. Even though mixed-effects modeling was the most popular approach, none of the articles specified the covariance structure used in the analysis.

Journal Name	Number of Relevant Articles	Year	Analysis Method Used
JREE	77	2008-2018	- ANCOVA (6, 7.8%) - (Fuzzy) Regression discontinuity (3, 3.9%) - Hierarchical linear model (HLM) (8, 10.4%) - Linear regression (7, 9.0%) - MANCOVA / MANOVA / multivariate linear regression / repeated-measures ANOVA (5, 6.5%) - Mixed-effects models (7, 9.1%) - Multilevel modeling (17, 22.1%) - SEM-related (7, 9.1%) - Time-series analysis (1, 1.3%) - Power analysis (2, 2.6%) - Propensity score (5, 6.5%) - t-test (1, 1.3%) - Others (9, 11.7%)
JAP	75	2009-2018	- Cox regression model (3, 4.0%) - Exponential random graph (ERG) model (2, 2.7%) - Hierarchical linear model (HLM) (5, 6.5%) - Linear regression (14, 18.7%) - Logistic regression (1, 1.3%) - Longitudinal Probit model (1, 1.3%) - Meta analysis (2, 2.7%) - Mixed-effects model (5, 6.7%) - Multilevel modeling (4, 5.3%) - Random coefficients model (4, 5.3%) - SEM-related (34, 45.3%)
DP	94	Randomly chosen articles in 2010-2019	- ANCOVA (2, 2.1%) - Hierarchical linear model (HLM) (4, 4.3%) - Linear regression (12, 12.8%) - MANOVA (4, 4.3%) - Mixed-effects models (3, 3.2%) - Multilevel modeling (5, 5.3%) - Repeated measures ANOVA (4, 4.3%) - SEM related (51, 54.3%) - Paired t-test (1, 1.1%) - Others (8, 11.7%)
EEPA	80	Randomly chosen articles in 2010-2019	- Difference-in-differences (DiD) analytical approach (5, 6.3%) - GEE (1, 1.3%) - Hierarchical linear model (HLM) (5, 6.3%) - Instrumental variables models (3, 3.8%) - Time-series analysis (2. 2.5%) - Logit model (i.e., Generalized Linear Mixed Model) (8, 10.0%) - MANOVA (1, 1.3%)

	- Mixed-effects model (16, 20.0%) - Multilevel modeling (5, 6.3%) - Ordinary least squares (OLS) regression (6, 7.5%) - Qualitative data analysis (2, 2.5%) - Regression discontinuity (5, 6.3%) - SEM-related (1, 1.3%) - Time-to-event data analysis (3, 3.8%) - Two-stage least squares approach (2SLS) (3, 3.8%) - Value-added model (4, 5.0%) - Others (10, 12.5%)

Table 1. *Journal survey results*

In all of the surveyed journals, the mixed-effects modeling approach was among the top two methods used. While it is encouraging that researchers are using more advanced methods, in all but a few cases the covariance structure of the repeated measure data was not reported. This could mean that scholars are simply implementing the default methods in the programming language and might not be considering the assumptions that are applied to the structure of the repeated measure data in the background.

As discussed in the next section, the same set of surveyed articles was used to identify the average sample size and number of time points in longitudinal studies in Education and Psychology.

2.1. Sample Size and Number of Repeated Measures for Reviewed Articles

Further review of sample size and number of time points (i.e., number of repeated measures) was conducted for the articles published in JREE, JAP, DP, and EEPA. This information can help researchers to identify the disciplines' norms for sample size and number of repeated measures. Identifying these norms is essential for several reasons, including when running simulation studies to test the properties of certain longitudinal models.

For JREE, articles that were systematic reviews (N = 3), meta-analyses (N = 2), simulation studies (N = 1), related to power analysis (N = 2) or not related to longitudinal studies (N = 13) were not considered in this further investigation. One article that did not indicate sample size for the archived data used, but did indicate the data sources, was also excluded in this investigation. The final number of articles in JREE reviewed for sample size and number of repeated measures was 68.

For JAP, articles that were meta analyses (N = 2), or not related to longitudinal studies (N = 7) were not considered in this further investigation. The final number of articles in JREE reviewed for sample size and number of repeated measure was 73.

For DP, articles that were not related to longitudinal studies (N = 6) were not considered in this further investigation. The final number of articles in DP reviewed for sample size and number of repeated measures was 94.

For EEPA, the following articles were not included in the further investigation for *sample size*:

- Articles that were not related to longitudinal studies (N = 20),
- One article that did not indicate sample size for the archived data used, but did indicate the data sources (N = 1),
- One article for meta-analysis (N = 1),
- One article that used four data sets to demonstrate formula for effect size computation (N = 1),
- Articles that only provided information for total number of observations for the entire study period (e.g. 10 subjects observed four times, so total number of observations = 40) (N = 9).

The final number of articles in EEPA reviewed for *sample size* was 68.

For EEPA, the following articles were not included in the further investigation for *number of time points*:

- Articles that were not related to longitudinal studies (N = 20),
- One article that did not specify number of time points used as "*because this project focuses on college choice, I limit my data to American citizens or permanent residents who were accepted to* ***at least two of the sampled colleges*** *in the spring of 2009*" (N = 1),
- One article that used four data sets to demonstrate formula for effect size computation (N = 1),
- One article using meta-analysis (N = 1).

The final number of articles in EEPA reviewed for *number of time points* was 77.

Table 2 presents the descriptive statistics of sample size and number of time points for articles reviewed in JREE, JAP, DP, and EEPA. For JREE, the sample size of the articles ranged from 44 to 1905147, with a median sample size of 1116; the number of time points for studies ranged from 2 to 15, with a median number of repeated measures equal to 2. For JAP, the sample size of the articles ranged from 20 to 49242, with a median sample size of 458; the number of time points for studies ranged from 2 to 48, with a median number of repeated measures equal to 3. For DP, the sample size of the articles ranged from 18 to 38017, with a median sample size of 541.5; the number of time points for studies ranged from 2 to 17, with a median number of repeated measures equal to 3. For EEPA, the sample size of the articles ranged from 30 to 4109265, with a median sample size equal to 5832.5; the number of time points for studies ranged from 2 to 33, with a median number of repeated measures equal to 4.

Table 3 shows the frequency distribution of sample size and number of repeated measures for articles reviewed in JREE, JAP, DP, and EEPA. For JREE, sample sizes from 1001-10000 were the most commonly adopted (29.41%) in the articles reviewed. For JAP and DP, sample sizes from 101-500 were the most commonly adopted (45.21%, 36.17%) in the articles reviewed. For EEPA, sample sizes of 10000+ were the most commonly adopted (39.71%) in the articles reviewed. For all the reviewed journals, the majority of the articles were studies with 2-5 time points (95.59% for JREE, 86.30% for JAP, 80.85% for DP, and 58.44% for EEPA).

Figures 1 to 4 present the bar graphs of the results from Table 3 for easier visual inspection. These figures suggest that scholars either are not collecting a sufficient number of time points (i.e. repeated measures) or they are using the longitudinal data partially. This means literature in Education and Psychology can benefit from collecting more repeated measurements of the same units and/or from using all available time points when analyzing data longitudinally. If either procedure is implemented, researchers could then use more advanced methods with better statistical properties. Finally, although there exist small sample sizes (i.e. smaller than 30), the majority of sample sizes are in an acceptable range.

Journal		Mean (SD)	Median (Range)	Min	Max
JREE	Sample size	36464.10 (230975.34)	1116 (1905103)	44	1905147
	Number of time points	3.13 (2.12)	2 (13)	2	15
JAP	Sample size	2050.33 (6342.26)	458 (49222)	20	49242
	Number of time points	4.41 (6.26)	3 (46)	2	48
DP	Sample size	1974.96 (5586.46)	541.5 (37999)	18	38017
	Number of time points	4.01 (2.36)	3 (15)	2	17
EEPA	Sample size	124092.10 (562955.10)	5832.5 (4109235)	30	4109265
	Number of time points	5.90 (5.29)	4 (31)	2	33

Table 2. *Descriptive statistics of sample size and number of repeated measures for articles reviewed in JREE, JAP, DP and EEPA*

		JREE	JAP	DP	EEPA
		$N_{SampleSize} = 68$	$N_{SampleSize} = 73$	$N_{SampleSize} = 94$	$N_{SampleSize} = 68$
		$N_{TimePoints} = 68$	$N_{TimePoints} = 73$	$N_{TimePoints} = 94$	$N_{TimePoints} = 77$
Sample size	1-100	7 (10.29%)	6 (8.22%)	10 (10.64%)	4 (5.88%)
	101-500	19 (27.94%)	33 (45.21%)	34 (36.17%)	4 (5.88%)
	501-1000	8 (11.76%)	16 (21.92%)	17 (18.09%)	9 (13.24%)
	1001-10000	20 (29.41%)	15 (20.55%)	30 (31.91%)	24 (35.29%)
	≥10000	14 (20.59%)	3 (4.11%)	3 (3.19%)	27 (39.71%)
Number of time points	2-5	65 (95.59%)	63 (86.30%)	76 (80.85%)	45 (58.44%)
	6-10	2 (2.94%)	6 (8.22%)	15 (15.96%)	22 (28.57%)
	≥10	1 (1.47%)	4 (5.48%)	3 (3.19%)	10 (12.99%)

Table 3. *Count (percentages) of sample size and number of repeated measures for articles reviewed in JREE, JAP, DP and EEPA*

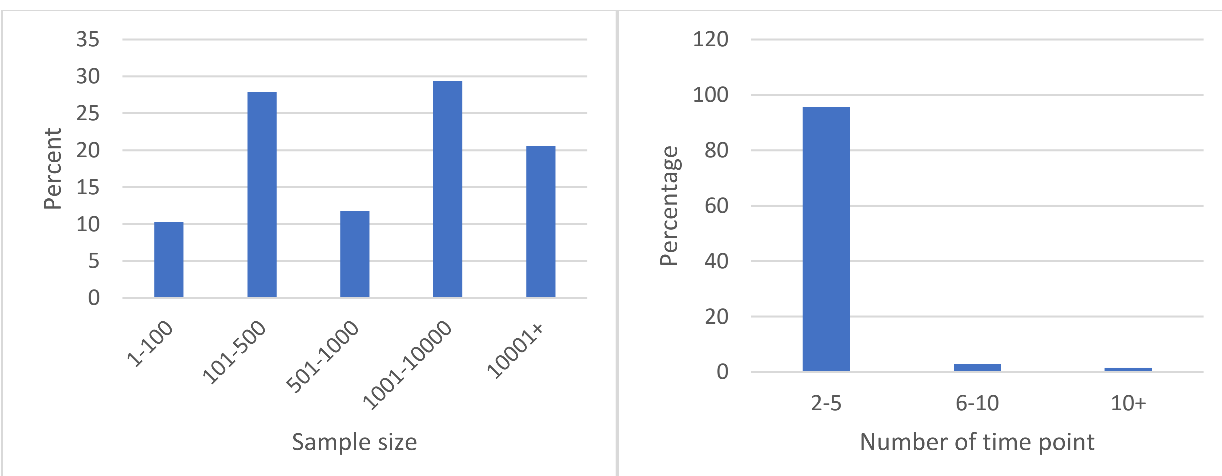


Figure 1. *% of sample size and % of number of repeated measures for JREE*

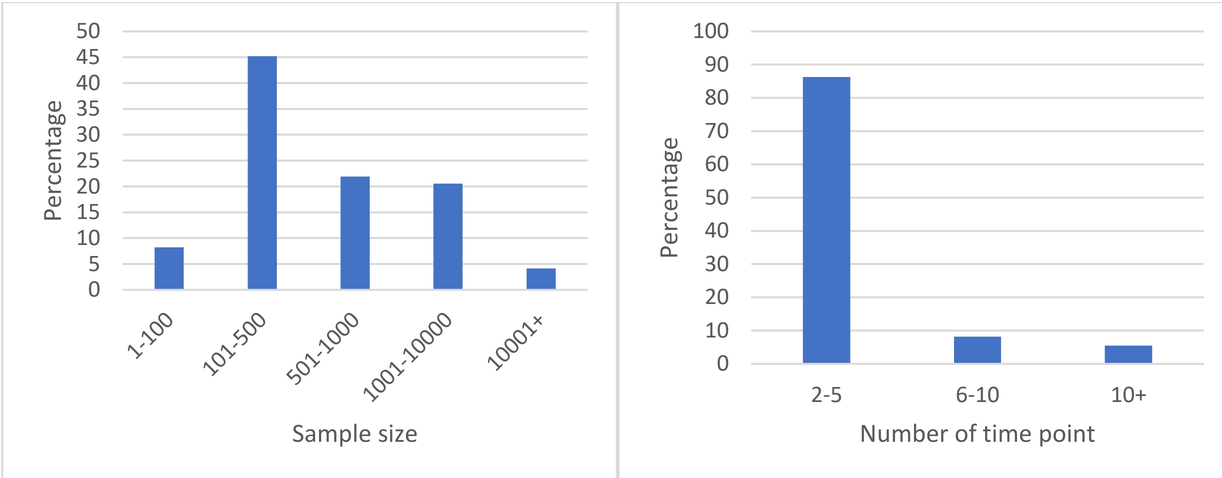


Figure 2. *% of sample size and % of number of repeated measures for JAP*

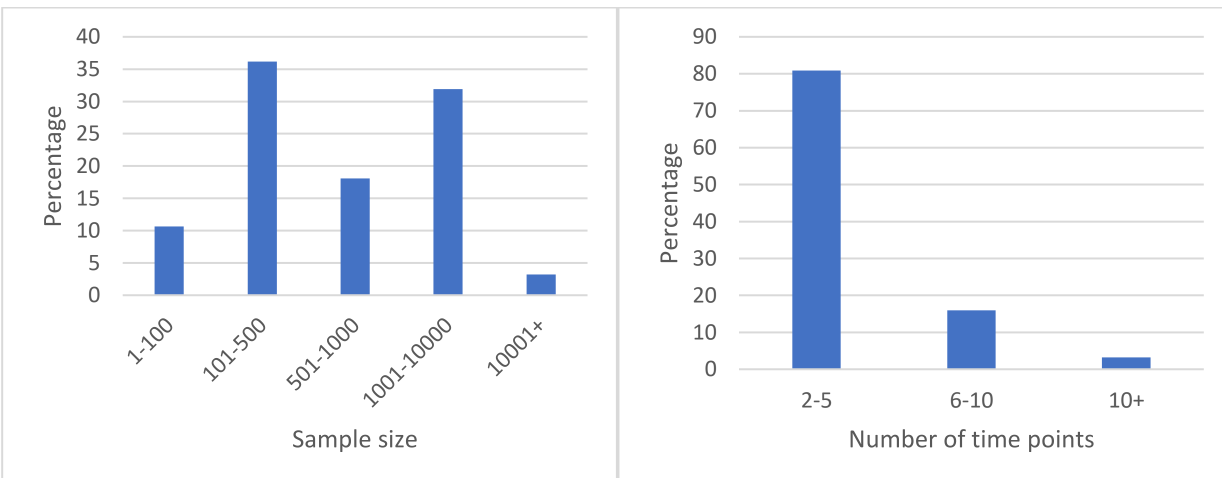


Figure 3. *% of sample size and % of number of repeated measures for DP*

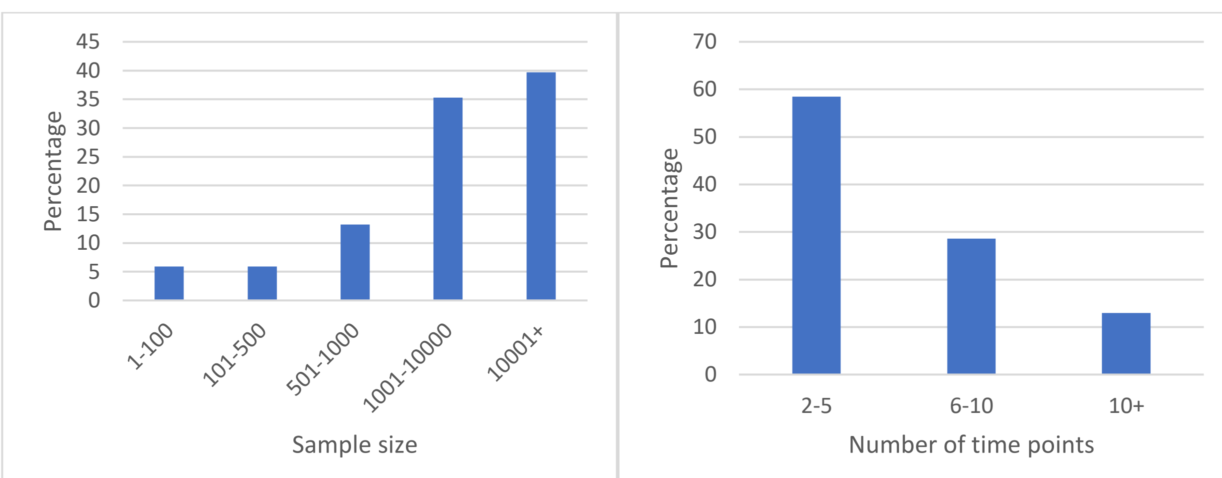


Figure 4. *% of sample size and % of number of repeated measures for EEPA*

3. Data Analysis Review and Methods for Longitudinal Data in Education and Psychology

While a survey of four journals indicates that sophisticated models such as Linear Mixed Modeling (LMM) techniques are commonly used, traditional models such as ANOVA continue to appear in the literature. Although traditional models are easy to understand and implement, they come with rigid assumptions that are unrealistic in most cases in Education and Psychology. The statistical details of these assumptions will be covered shortly.

This section reviews longitudinal models in Education and Psychology and provides references for scholars who seek further examples. Following each model's review section, there follows a concise methods section that can help researchers to identify the statistical strengths and weaknesses of each method.

All methods are categorized as either "traditional" or "advanced." Traditional methods include methods such as ANOVA and simple linear regression (which generally analyzes longitudinal data cross-sectionally), among many others. Advanced methods include LMMs and GEEs, among many others.

3.1. Traditional Approaches

The review and methods details of traditional approaches will be discussed in this section. Traditional approaches to analyzing longitudinal data include the paired t-test, analysis of variance (ANOVA), analysis of covariance (ANCOVA), regression analysis, derived variable approach, and repeated measures univariate/multivariate analysis of variance (RM ANOVA and RM MANOVA). Note that the derived variable approach is here considered as a subtopic of each analytical method.

In a study with two time points such as pretest and posttest, the main purpose may be to determine whether:

1. The outcome variable changes significantly over time,
2. The posttest score is related to the intervention after controlling for the pretest score,
3. The change over time is associated with the intervention and the fixed features of subjects (Twisk, 2003).

These three objectives will be addressed sequentially in the following sections.

3.1.1. Review: Paired t-test

This first objective mentioned in section 3.1. could be addressed utilizing the paired t-test with time as the independent factor. For instance, using data on bullying collected from March to April, 2012 (i.e. Time 1) and from March to April, 2013 (i.e. Time 2), Hellfeldt, Gill, and Johansson (2018) utilized four separate paired t-tests made for four victimization profiles during the study period:

1. Never bullied subjects throughout the measurement period were defined as "non-victims,"
2. Subjects whose status changed from victim to non-victim (from Time 1 to Time 2) were defined as "ceased victims,"
3. Subjects whose status changed from non-victim to victim (from Time 1 to Time 2) were defined as "new victims,"
4. Subjects who were bullied at both measurement points were defined as "persistent or continuing victims."

The researchers used these paired t-tests to determine changes in psychological well-being for the four types of bullying victims among pupils from 44 elementary schools, from 4th to 9th grade, in a medium-sized Swedish city. More examples of the applications of the paired t-test on longitudinal data in the field of Education and Psychology can be found in Bartl, Hagl, Kotoučová, Pfoh, and Rosner (2018); Bensley, Crowe, Bernhardt, Buckner, and Allman (2010); Hwang and Chang (2011); Konishi, Hymel, Danbrook, and Wong (2018); Martin and Calvert (2018); and Pittman and Richmond (2008).

The paired t-test can in fact be seen as one of the simplest longitudinal analysis methods as it represents the case of a single group of subjects, each of which has been measured on two occasions; it can be used to measure whether there has been significant average change between the two time points.

3.1.1.1. Method: Paired t-test

The paired t-test is a statistical method used to examine the equality of the means of two sets of related or matched observations, or (what amounts to the same thing) assessing whether the observed difference in mean between the two sets of values is zero. This method is also called the dependent sample t-test.

Let us assume our data consist of N participants. Pre- and post-test variables measured for participant i are shown as y_{i1} and y_{i2}, respectively. Then $d_i = y_{i2} - y_{i1}$ is the difference between the two measures, or *change score*, for subject i.

Hypotheses and test statistic - The null hypothesis assumes that the true mean difference between the pre-test and post-test measurements is zero and can be written as:

$$H_0\colon \mu_1 = \mu_2 \text{ or, equivalently as } H_0\colon (\mu_2 - \mu_1) = \mu_d = 0$$

The alternative hypothesis can be written in a few different ways depending on the question of interest, as follows:

$$H_1\colon \mu_1 \neq \mu_2 \text{ or } H_1\colon \mu_d \neq 0 \text{ (two-tailed)}$$

$$H_1\colon \mu_2 > \mu_1 \text{ or } H_1\colon \mu_d > 0 \text{ (upper-tailed)}$$

$$H_1 : \mu_2 < \mu_1 \text{ or } H_1 : \mu_d < 0 \text{ (lower-tailed)}$$

The test statistic is calculated as:

$$t = \frac{\bar{d}}{\left(\frac{s_d}{\sqrt{N}}\right)} = \frac{\bar{d}}{\left(\frac{\sqrt{\frac{\left[\sum_i d_i^2 - \frac{(\sum_i d_i)^2}{N}\right]}{N-1}}}{\sqrt{N}}\right)} \sim t_{N-1}$$

which, under the null hypothesis, follows a Student's $t(N-1)$, where $\bar{d} = \frac{\sum_i d_i}{N}$.

Although normal distribution of the response is assumed in a paired t-test, it is fairly robust to departures from the normality assumption. Note that in this setting the outcome variable should be continuous (interval/ratio). The paired t-test is equivalent to conducting a simple linear regression where the change score is the outcome variable $d_i = \beta_0 + e_i$ and testing $H_0 : \beta_0 = 0$, whose corresponding statistic is $\frac{\widehat{\beta}_0}{se(\widehat{\beta}_0)}$, which follows a Student's $t(N-1)$.

3.1.2. Review: Analysis of Covariance (ANCOVA)

The second objective mentioned in section 3.1—determining whether posttest score is related to the intervention after controlling for the pretest score—is typically addressed using ANCOVA. An ANCOVA can be conducted to measure the effects of an experiment on the variables of interest. This approach both provides higher power and handles the effects of pre-test scores in the assessment of the differences between treatment groups when evaluating change resulting from formal interventions (Dimitrov & Rumrill, 2003). In ANCOVA, the pre-test score is used as a covariate. These analyses partial out the pre-test scores and then examine differences between the groups on the post-test. The study conducted by Piro and Ortiz (2009) used ANCOVA to explore "the effects of a scaffolded music instruction program on the vocabulary and verbal sequencing skills of two cohorts of second-grade students" from two public elementary schools in the same middle-class area of New York City. During the study period, the experimental group of $N = 46$ studied piano formally for 3 successive years as part of the experiment, and the control group of $N = 57$ had no experience with music lessons. It was shown that "the experimental group had significantly better vocabulary and verbal sequencing scores at post-test than did the control group" (Piro & Ortiz, 2009). Other examples of more recent studies that also employed ANCOVA to examine the effect of an intervention on post-test measure while adjusting for pre-test measure include: Hwang and Chang (2011); Vos, van der Meijden, and Denessen (2011); Uhls et al. (2014); Hermanto and Zuroff (2018); Bartl et al. (2018); and Martin and Calvert (2018).

3.1.2.1. Method: Analysis of Covariance (ANCOVA)

The ANCOVA can be seen as an extension of an ANOVA (which will be covered shortly) where a continuous variable (sometimes called covariates) has been added to the model. It is equivalent to a multiple regression (when there are no repeated measures). When there are repeated measures, in the simplest case, the ANCOVA is equivalent to a LMM.

In the context of longitudinal data, it is important to consider that a covariate can be time variant (i.e. it varies across subject and time points) or time invariant (i.e. the covariate values are the same across time for a given subject). An example of a time variant characteristic would be salary and of a time invariant covariate gender.

One of the simplest examples of an ANCOVA is the model for the post-test scores. In this model, there are two repeated measures per subject, pre- and post-test scores, but the post-test is used as a response variable and the pre-test is used as a covariate. This model can be written as:

$$y_{i2} = \beta_0 + \beta_1 x_i + \beta_2 y_{i1} + e_i$$

where x_i is the dummy variable for treatment (this model assumes only two groups: control vs. treatment), y_{i1} represents the pre-test scores and y_{i2} the post-test scores. The focus of this model is testing $H_0 \colon \beta_1 = 0$ (i.e. whether the mean of post-test is the same for both groups, after controlling for the pre-test).

The multiple group RM ANCOVA with one covariate and two groups can be written as:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 t_{2j} + \beta_3 t_{3j} + \beta_4 x_{2ij} + \pi_i + e_{ij}$$

where,

- $i = 1, \dots, N$ corresponds to subject index,
- $j = 1, \dots, n$ corresponds to measurement index (note, $n = 3$),
- y_{ij} denotes the outcome for subject i at time j,
- x_{1i} denotes the dummy for treatment (which is equal to 1 if subject i belongs to treatment and equals 0 if subject i belongs to control),
- t_{2j} represents the dummy for time = 2 (which is equal to 1 for all observations measured at time point 2 ($j = 2$) and 0 elsewhere),
- t_{3j} represents the dummy for time = 3 (which is equal to 1 for all observations measured at time point 3 ($j = 3$) and 0 elsewhere),
- x_{2ij} represents the value of the time variant variable for subject i at time j,
- π_i denotes the subject-specific component (i.e. random intercept),
- e_{ij} denotes the error term for subject i at time j.

Note that, for the sake of simplicity, this model assumes no interaction between time and group.

The assumptions for the ANCOVA model include all those for the RM ANOVA, which will be covered shortly, with the additional assumptions that the relationship between y and the variable is linear and the slope (between variable and response) is equal across groups.

3.1.3. Review: Analysis of Variance (ANOVA)

The third objective mentioned in section 3.1—determining whether change over time is associated with the intervention and the fixed features of subjects—can be addressed using approaches such as ANOVA and regression analysis with the change score (e.g.: posttest score – pretest score) as the outcome, and the variable(s) of interest (e.g.: intervention and demographic factors) as the independent variable(s). Change scores offer an unbiased estimate of true change irrespective of baseline value, and analysis using change scores is considered statistically similar to a repeated measures analysis (Zumbo, 1999). In a study conducted by Schonert-Reichl and Lawlor (2010), "pre- and early adolescent students in the 4th to 7th grades (N=246) drawn from six [Mindfulness Education] ME program classrooms and six comparison classrooms (wait-list controls) completed pretest and posttest self-report measures assessing optimism, general and school self-concept, and positive and negative affect." The researchers were interested in exploring the direction of change in students' "well-being and social and emotional competence" from pretest to posttest, and hence a series of ANOVAs were conducted using change score (computed as posttest score minus pretest score using the self-report measures) as the outcome, group (ME program vs. Control) as the independent variable, and students' gender, age, and first language learned as control variables (Schonert-Reichl & Lawlor, 2010). The analysis results revealed that pre- and early youths who participated in the ME program had significant increases in optimism from pretest to posttest compared to those who did not participate (Schonert-Reichl & Lawlor, 2010). A similar application of ANOVA can be seen in Bensley et al. (2010). Statistical details of this type of ANOVA are similar to those of ANCOVA and will not be covered independently here. However, all details related to the four RM ANOVA methods (i.e. univariate or multivariate, and with single or multiple sample) will be covered shortly.

3.1.4. Review: Regression Analysis

Because change scores (i.e. the dependent variables) are inclined to have greater measurement error and lower reliability compared to the original measurement scores (Allison, 1990; Zumbo, 1999), a slightly different form of change score analysis was used by some researchers. For example, Konishi et al. (2018) estimated residualized difference scores using the regression analysis (i.e. ordinary least square) of Time 2 on Time 1 for all of the outcome variables (i.e., number of friends inside and outside school separately, competitiveness, self-worth, and bullying). The residualized difference scores then served as the outcome variables of the regression models to study changes in bullying behavior in relation to friends, competitiveness, and self-worth among students in Grades 5 to 7 in Canada (Konishi et al., 2018). They found that children's beliefs about their self-worth were vital in predicting changes in bullying behavior where increased self-worth was associated with a decrease in reported bullying behavior (Konishi et al., 2018). Similar application of the residualized difference scores can also be seen in Pittman and Richmond, (2008) and Rubin, Evans, and Wilkinson (2016). Since regression analysis is a very broad umbrella, methods related to regression approaches for repeated measure data will be covered in detail in several different sections below.

3.1.5. Review: Derived Variable Approach

The derived variable approach reduces the repeated measurements into a summary variable (Hedeker & Gibbons, 2006). That is, given a vector of observations on a particular subject, a derived variable is a scalar-valued function of the vector of observations (Diggle et al., 2002). According to Hedeker and Gibbons (2006), examples of the derived variables approach for longitudinal data include but are not limited to:

- Carrying the last observation forward,

- Change score,
- Average over time,
- Linear trend over time,
- Area under the curve.

A key motivation for applying the derived variable approach on longitudinal data is that standard methods, such as 2-sample t-test, ANOVA, and regression analysis, can be used for inference (Diggle et al., 2002; Hedeker & Gibbons, 2006). Though convenient, this approach has several limitations (Diggle et al., 2002; Hedeker & Gibbons, 2006). For example, this approach is not applicable if there are any incomplete data, since individuals with incomplete data will need to be omitted or other missing data methods must be used. Also, uncertainty in the derived variable approach is proportional to the number of measurement occasions. When attrition or dropout (i.e. unbalanced data) leads different units to have different numbers of observations, different uncertainties arise. This means the homoscedasticity assumption is violated. Additionally, collapsing multiple repeated measurements to a single summary statistic may result in lower statistical power. Finally, due to removing the temporal aspect of the data, it is not possible to include time-varying variables. Regardless, the derived variable approach has been used in Education and Psychology longitudinal research. See Rapport et al. (2008), Oxford and Lee (2011), and Russell, Lee, Spieker, and Oxford (2016) for examples of the application of this approach.

3.1.5.1. Method: Derived Variable Approach

One of the simplest methods of treating longitudinal data is called the derived variable approach. As mentioned before, this approach reduces the repeated measures into a single summary variable. This summary variable can be the average across time, linear trend, change score, or area under the curve. This method transforms the data into independent observations where there will be a single observation, the summary measurement, per individual. In this way, the traditional non-longitudinal statistical methods can be applied to the transformed data.

The following example demonstrates the case in which there are two repeated measures per subject and the summary measurement is the difference between the pre and post values of the dependent variable, also known as *change score*. The regression model for the change score can be presented as:

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

$$e_i \sim N(0, \sigma^2)$$

where y_i is defined as the difference between the occasions for subject i, (i.e. $y_i = y_{i2} - y_{i1}$), x_i equals 1 for the treatment group and 0 for the control group.

Note that this model characterizes an ordinary regression model so it is subject to the assumptions of the linear regression model where additional covariates could be easily added to the model. Note that the disadvantages of the derived variable approach mentioned in the previous section still stand.

3.1.6. Review: Repeated Measures Univariate and Multivariate Analysis of Variance (RM ANOVA and RM MANOVA)

When outcome variables are collected at two or more time points on the same subjects, RM ANOVA and RM MANOVA can be used to compare the means of the time points. These methods can be utilized to evaluate whether the outcome has changed significantly across time points. Yet neither method provides information about subject-specific pattern over time (Newsom, 2012). Questions often asked in analyses using RM ANOVA and RM MANOVA include:

1. Is there a difference in the dependent variable between the groups, regardless of time?

2. Is there a difference in the dependent variable between different time points, regardless of groups?
3. Does the difference in the dependent variable between groups vary over time? (Newsom, 2012).

The use of RM ANOVA and RM MANOVA are restricted due to the limiting missing data assumptions across time and the specific covariance pattern of the time points (Hedeker & Gibbons 2006). RM ANOVA requires all subjects to be measured the same number of time points, and RM MANOVA allows no missingness. Furthermore, a disadvantage of RM ANOVA is the assumption that the outcome measures have equal variances and covariances over time (i.e., compound symmetry), which might be unrealistic because most of the time variance increases with time and covariance decreases with increasing time lags. On the other hand, the RM MANOVA model imposes no assumptions on the variance-covariance structure of the repeated measurements.

RM ANOVA and RM MANOVA have been widely used to analyze longitudinal data in the fields of Education and Psychology. For example, Fuchs, Compton, Fuchs, Bryant, and Davis (2008) employed RM ANOVA to explore differences in groups for measures of reading at pre-test, mid-year, and post-test for data collected from 252 first-grade children in middle Tennessee. The individuals were randomly assigned into the following tutoring groups, each $n = 84$:

- Fall tutoring group: All students were part of small-group tutoring during the fall semester for 9 weeks;
- Spring tutoring group: Participants non-responsive to fall semester training were assigned to small-group tutoring during the spring semester for 9 weeks;
- Control group: Students were matched to the non-responding participants in the spring tutoring group.

Lee and Zentall (2015) conducted a 3-year longitudinal study to investigate reading motivation and achievement. The authors used the RM MANOVA to assess the between-group factor of disability (reading disabilities, attention deficit hyperactivity disorder, and no disabilities) and the within-group time factor (elementary to middle school levels) for outcomes such as self-efficacy, social motivation, and work avoidance. Other examples of RM MANOVA applications in Education and Psychology include: Cemalcilar and Falbo (2008) and Myers (2017). Other examples of RM ANOVA applications in Education and Psychology include Blonigen, Carlson, Hicks, Krueger and Iacono (2008); Kim et al. (2015); Breeman, Jaekel, Baumann, Bartmann and Wolke (2016); Aelterman, Vansteenkiste, Van Keer and Haerens (2016).

3.1.6.1. Methods: RM ANOVA and RM MANOVA

Although more advanced statistical techniques (such as multilevel or mixed-model analyses) now exist that can better analyze longitudinal data, the ANOVA offers two classical approaches to longitudinal data analysis: the repeated measures ANOVA (RM ANOVA) and the multivariate ANOVA (RM MANOVA or just MANOVA). These models are worth reviewing to set up a basis for understanding more advanced methods.

The limitations of these two ANOVA approaches are that the repeated measures are assumed to be fixed across subjects (i.e. time 1 for subject 1 needs to be the same as time 1 for subject 2 and so on). An example of a longitudinal study where time points are not “fixed occasions” is when students take computerized exams in a certain month on a first-come, first-served basis. Another limitation shared by both methods is the use of least squares estimation, which makes them more vulnerable to the presence of outliers and missing data. More specifically, although the RM ANOVA can be used with unbalanced data (i.e. missing data), the MANOVA cannot handle missing data, forcing the researcher to delete all incomplete cases from the study. Needless to say, this can introduce an unwanted bias to the model estimates.

3.1.6.1.1. Method: Single-Sample RM ANOVA

This model represents one of the simplest repeated measures design, where there is only one sample of participants measured over time and *no groups of subjects* being compared (e.g. intervention vs. control groups).

Let y denote the dependent variable, $i = 1, \ldots, N$ corresponds to subject index and $j = 1, \ldots, n$ is the time index or occasions. The model can be written as:

$$y_{ij} = \mu + \pi_i + \tau_j + e_{ij},$$

$$\pi_i \sim N(0, \sigma_\pi^2)$$

$$e_{ij}|\tau_j \sim N(0, \sigma_e^2)$$

where,

- μ denotes the overall mean or model intercept,
- π_i denotes the subject-specific deviation from the overall mean (i.e. random intercept),
- τ_j denotes the time effect, assumed the same for all subjects,
- e_{ij} denotes the error corresponding to subject i measured on occasion j,
- σ_π^2 denotes the between-subject variance,
- σ_e^2 denotes the within-subject variance.

The subject-specific error term π_i remains constant across time points for a single individual, while the second error term or residual e_{ij} is occasion-specific and varies between subjects i and occasions j .

Model assumptions – are as follows:

1. Normality of the outcome variable: This assumption is an extension of the assumption that the e_{ij}s are normally distributed within each level of the within-subject factor (time point). The RM ANOVA is quite robust to the violation of the normality assumption.
2. Sum of the n time parameters is constrained to be zero (i.e. $\sum_{j=1}^{n} \tau_j = 0$).
3. The random components and the error terms both have a zero mean. From this assumption it follows that the expectation of y_{ij} is the grand mean plus the time effect (i.e. $E(y_{ij}) = \mu + \tau_j$).
4. The random components π_i are independent of the residuals e_{ij}. From this assumption it follows that $Var(y_{ij}) = Var(\mu + \pi_i + \tau_j + e_{ij}) = \sigma_\pi^2 + \sigma_e^2$.
5. Independence between subjects are also assumed which translates to $Cov(y_{ij}, y_{i'i}) = 0 \ for \ i \neq i'$.
6. Constant covariance between observations within the same subject (i.e. $Cov(y_{ij}, y_{ij'}) = \sigma_\pi^2 \ for \ j \neq j'$).

E(.), Var(.) and Cov(.) represent the expectation, variance, and covariance functions, respectively.

Assumptions 5 and 6 induces the following variance-covariance pattern:

$$\begin{bmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \cdots & \sigma_\pi^2 & \sigma_\pi^2 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \vdots & & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 \end{bmatrix}$$

This structure is known as *compound symmetry (CS) or exchangeable*. The variance is homogeneous or constant across time, represented by the diagonal of terms $\sigma_\pi^2 + \sigma_e^2$ and the covariances are homogeneous across time, represented by σ_π^2. As mentioned before, this assumption is not very realistic since one would expect that variances change over time and covariances of responses closer in time should be more correlated, compared to responses that are more distant in time.

Intra-class correlation (ICC) - The covariance defined under assumption 6 can be written in the form of the following correlation:

$$Cor(y_{ij}, y_{ij'}) = \frac{Cov(y_{ij}, y_{ij'})}{\sqrt{Var(y_{ij})}\sqrt{Var(y_{ij'})}} = \frac{\sigma_\pi^2}{\sqrt{\sigma_\pi^2 + \sigma_e^2}\sqrt{\sigma_\pi^2 + \sigma_e^2}} = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_e^2}$$

The ICC represents the magnitude of the within-subject correlation or the between-subject heterogeneity. Since the elements of both numerator and denominator are variances (i.e. always positive), then the ICC ranges from 0 to 1. The extreme case of ICC being equal to 0 happens when there is no between-subject variance (i.e. $\sigma_\pi^2 = 0$) and the case in which ICC equals 1 is when the between-subject variance explains all the variance, in other words, there is no heterogeneity in the repeated measures of the same subject. The ICC can be defined as the proportion of unexplained variation that is due to subjects.

Sphericity - CS is a sufficient assumption to ensure that the F-test of the RM ANOVA follows an F distribution but is not necessary. Sphericity or circularity is a less restrictive assumption imposed on the structure of the covariance matrix compared to the CS assumption; it is a sufficient and necessary condition of the RM ANOVA. *Sphericity* is defined as the equality of all the variances of the differences between any two levels of the within-subject factor (i.e. time points). Note that sphericity only has a meaning when there are more than two levels of the within-subjects factor or time points.

$$Var(y_{ij}, y_{ij'}) = Var(y_{ij}) + Var(y_{ij'}) - 2Cov(y_{ij}, y_{ij'}) = constant, \qquad \forall\, j \ and \ j'$$

Note that if the CS condition is met, it implies that the sphericity condition is satisfied.

The chi-square goodness-of-fit test developed by Mauchly (1940), known as Mauchly's sphericity test, is generally used to test for sphericity. Researcher caution is required since this test is not very reliable for small samples and tends to be too sensitive for large samples (i.e., it may show significance even when there is a minor departure from sphericity). This test also is sensitive to the presence of outliers and deviations from normality. Knowing all the disadvantages of Mauchly's test, researchers should not use it as a strict rule but as a guide.

Solutions to the violation of sphericity – Alternatively, when the sphericity assumption is rejected, the use of adjusted p-values for the F-tests is recommended. These corrections were developed by Greenhouse and Geisser (1959) and Huynh and Feldt (1976). Both corrections work in a very similar way and tend to be very conservative.

Another solution is the use of the multivariate repeated measures analysis, which allows a more general structure of the variance-covariance matrix (i.e., it does not assume sphericity). However, recall that MANOVA can only work with complete data across time.

Table 4 introduces the ANOVA table corresponding to a balanced design to be used for testing; it is also a good baseline for future sections.

Source	**df**	**SS**	**MS**
Subjects	$N - 1$	$SS_S = n\sum_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{SS_S}{N-1}$
Time	$n - 1$	$SS_T = N\sum_{i=1}^{N}(\bar{y}_{.j} - \bar{y}_{..})^2$	$\frac{SS_T}{n-1}$
Residual	$(N - 1) \times (n - 1)$	$SS_R = \sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$\frac{SS_R}{(N-1)(n-1)}$
Total	$Nn - 1$	$SS_y = \sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	

Table 4. ***"****SS" stands for sum of squares, "MS" stands for mean squares, "$\bar{y}_{..}$" represents the overall mean, "$\bar{y}_{i.}$" the mean for subject i, and "$\bar{y}_{.j}$" the mean for time point j*

Hypothesis testing – For this simple model, there are only two types of tests, namely, testing time and subject effect. The focus of the model will be testing the significance of the time effect which means testing whether there is a trend over time for the response variable.

Testing for Subject effect – The null hypothesis for testing the subject-specific effect can be written as:

$$H_0: \sigma_\pi^2 = 0$$

This is the test of whether there is significant variance due to differences between subjects. The statistic corresponding to the subject effect, $F_S = \frac{MS_S}{MS_R}$, follows an $F(N-1, (N-1)(n-1))$.

Testing for Time effect – The omnibus test of no difference over time is as follows:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_n = 0$$

The corresponding test statistic, $F_T = \frac{MS_T}{MS_R}$, follows a $F(n-1, (N-1)(n-1))$.

Commonly used contrasts for time – Let us define a set of $n-1$ contrasts $L_{j'}$ as:

$$L_{j'} = \sum_{j=1}^{T} c_{j'j} \bar{y}_{.j}, \quad j' = 1, \cdots n-1$$

where $\bar{y}_{.j}$ represents the time-point mean and $c_{j'j}$ represents the contrast coefficients. For any given contrast $L_{j'}$, the sum of the contrast coefficient must be 0 across the total number of occasions or time points (i.e. $\sum_{j=1}^{T} c_{j'j} = 0$).

The statistic used to test a contrast with null hypothesis $L_{j'} = 0$ is defined as:

$$F_{j'} = \frac{MS_{j'}}{MS_R} \sim F_{1,(N-1)(n-1)}$$

where $MS_{j'} = SS_{j'} = \frac{NL_{j'}^2}{\sum_{j=1}^{T} c_{j'j}^2}$ and $MS_R = \frac{SS_R}{(N-1)(n-1)} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{T}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}{(N-1)(n-1)}$.

Assuming time can be decomposed in $n-1$ independent sets of contrasts, also called orthogonal contrasts, then the sum of squares for time can be written as the contrasts sum of squares as follows:

$$SS_{time} = \sum_{j=1}^{T-1} SS_j$$

Depending on the different partitions of time one is interested in testing, different types of contrasts, expressed by their corresponding coefficients $c_{j'j}$, can be used. A brief review of the most commonly used contrasts is as follows:

Trend analysis – This type of contrast expresses the $n-1$ partitions as orthogonal polynomials. For instance, if one assumes that the model has $n = 4$ time points, then the contrast matrix is

$$C = \begin{bmatrix} -3/\sqrt{20} & -1/\sqrt{20} & 1/\sqrt{20} & 3/\sqrt{20} \\ 1/\sqrt{4} & -1/\sqrt{4} & -1/\sqrt{4} & 1/\sqrt{4} \\ -1/\sqrt{20} & 3/\sqrt{20} & -3/\sqrt{20} & 1/\sqrt{20} \end{bmatrix}$$

with the first row of the matrix representing the linear trend contrast, and the second and third rows representing the contrasts for quadratic and cubic trends. Note that the above matrix assumes that the time points are equally spaced.

Change relative to baseline – This is the contrast of any significant change over time compared to baseline, as measured by testing the difference between each time point and the first time point. The corresponding matrix for a model where $n = 4$ (i.e. 4 time points) is as follows:

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

The four time points can be presented as *T1*, *T2*, *T3*, and *T4*; the contrasts in the rows represent the difference between T2 and T1 (baseline), between T3 and T1 and between T4 and T1, respectively. Note that the reference point in this example is the first time point. If, for instance, the researcher would be interested in using the last time point as a reference, then the first columns of -1 would move to column 4 and columns 1, 2, and 3 would each move one position to the left. Also, note that this is not a set of orthogonal contrasts.

Consecutive time comparisons – This is a useful contrast if one is interested in knowing whether the outcome at each time point is significantly different from the outcome at the immediately previous time point. The matrix corresponding to this type of contrasts for $n = 4$, which are sometimes called profile contrasts, is as follows:

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

where the rows test the difference between *T*1 and *T2*, the difference between *T2* and *T3*, and the difference between *T3* and *T4*. Also, note that this is not a set of orthogonal contrasts.

Contrasting each time point to the mean of the subsequent time points – Also known as Helmert contrasts, here the researcher is interested in comparing each time point to the average of all subsequent time points. The matrix for a $n = 4$ model is as follows:

$$C = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

where the rows compare *T1* to the mean of *T2*, *T3*, and *T4*; *T2* to the mean of *T3* and *T4*; and *T3* versus *T4*. Note that Helmert contrasts are orthogonal.

Correction to multiple comparisons – Making multiple comparisons results in inflating the Type I error (or α level) or, in other words, increasing the probability of rejecting a true null hypothesis. The Bonferroni correction offers a relatively conservative way of adjusting the α level. The adjusted α level consists of dividing the original α level by the number of contrasts (i.e. $\alpha^* =$

$\frac{\alpha}{n-1}$). For instance, following previous examples, with three multiple comparisons, the original α level of .05 will become $\alpha^* = \frac{0.05}{3} = 0.017$. There are other corrections to multiple comparisons such as Scheffe, Sidak, or Tukey's method.

A less conservative alternative is to first test the overall test $H_0\colon \tau_1 = \tau_2 = \cdots = \tau_T$, and if this is rejected, then to test each individual contrast using the uncorrected α level.

3.1.6.1.2. Method: Multiple-Sample RM ANOVA

This model is more commonly used than the previous one because it incorporates different groups of subjects that the researcher has an interest in comparing. This statistical design is commonly used in randomized-controlled clinical trials, where participants are randomly assigned to various experimental groups and their outcomes are tracked over time.

Model assumptions – Let us assume there exist $h = 1, \cdots, s$ groups with $i = 1, \ldots, N_h$ subjects in group h, and $j = 1, \cdots, n$ time points. The total sample size N is the sum of the sample sizes for each group ($N = \sum_{h=1}^{s} N_h$). The model is written as:

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij},$$

$$\pi_{i(h)} \sim N(0, \sigma_\pi^2)$$

$$e_{hij} \sim N(0, \sigma_e^2)$$

where,

- y_{hij} denotes the observation for individual i in group h at time j,
- μ denotes the overall mean or model intercept,
- γ_h denotes the effect of group h, with constraint $\sum_{h=1}^{s} \gamma_h = 0$,
- τ_j denotes the time effect, with constraint $\sum_{j=1}^{n} \tau_j = 0$,

- $(\gamma\tau)_{hj}$ denotes the interaction effect between time j and group h, with constraint $\sum_h \sum_j (\gamma\tau)_{hj} = 0$,
- $\pi_{i(h)}$ denotes the subject-specific deviation component for participant i nested in group h (subjects are considered random effects),
- e_{hij} denotes the error term for subject i in group h measured at time j.

Note that the assumptions on the distribution of the error terms ($\pi_{i(h)}$ and e_{hij}), which leads to the CS for V(.), are the same as in the previous model (i.e. single-sample RM ANOVA). The definitions of sphericity and intra-class correlation are likewise the same as in the previous model. In this model, the design is assumed to be balanced with respect to the number of repeated measures per subject. The focus of the model will be testing the significance of the time effect, i.e. testing whether there is a trend over time in the outcome.

Hypothesis testing: Testing for Group by Time Interaction - The most important test in this model is the one corresponding to the interaction term between group and time as it will determine whether the differences between groups are not equal across time (i.e. H_0: $(\gamma\tau)_{11} = \cdots = (\gamma\tau)_{sn}$). In other words, this test will determine whether the between-group trend lines across time are parallel or whether one treatment was more effective than others. The statistics for this test can be expressed as:

$$F_{GroupxTime} = \frac{\frac{SS_{GT}}{(s-1)(n-1)}}{\frac{SS_R}{(N-s)(n-1)}} \sim F_{(s-1)(n-1),(N-s)(n-1)}$$

where SS_{GT} represents the Sum of Squares for the group by time interaction,

$$SS_{GT} = \sum_{h=1}^{s} \sum_{j=1}^{n} N_h (\bar{y}_{h.j} - \bar{y}_{h..} - \bar{y}_{..j} + \bar{y}_{...})^2$$

and SS_R the Sum of Squares for the residuals,

$$SS_R = \sum_{h=1}^{s}\sum_{i=1}^{N_h}\sum_{j=1}^{n}(y_{hij} - \bar{y}_{h.j} - \bar{y}_{hi.} + \bar{y}_{h..})^2.$$

Note that the dot in the subscript is representative of the unit being averaged. If the H_0 is rejected, then one can conclude that there is no single overall group effect because it differs over time. Additionally, there is no single overall time effect since it varies across groups. However, if H_0 cannot be rejected, the following main effects tests should be conducted:

Hypothesis testing: Testing for Time effect - As in the previous model, this is the overall test with null hypothesis being no difference over time, expressed as follows:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_n = 0$$

with,

$$F_{Time} = \frac{\frac{SS_T}{n-1}}{\frac{SS_R}{(N-s)(n-1)}} \sim F_{n-1,(N-s)(n-1)}$$

where SS_T represents the Sum of Squares for time, $SS_T = N\sum_{j=1}^{n}(\bar{y}_{..j} - \bar{y}_{...})^2$.

Hypothesis testing: Testing for Group effect – The null hypothesis is that there is no group effect, expressed as follows:

$$H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_s = 0,$$

and the statistic is written as:

$$F_{Group} = \frac{SS_G/(s-1)}{SS_{S(G)}/(N-s)} \sim F_{n-1,(N-s)}$$

where SS_G represents the Sum of Square for group, $SS_G = n\sum_{h=1}^{s} N_h(\bar{y}_{h..} - \bar{y}_{...})^2$ and $SS_{S(G)}$ denotes the Sum of Squares for subjects in groups, $SS_{S(G)} = n\sum_{h=1}^{s}\sum_{i=1}^{N_h}(\bar{y}_{hi.} - \bar{y}_{h..})^2$.

Hypothesis testing: Testing for subject effect – This is the test of whether the random subject effects are different from zero. The null hypothesis is defined as:

$$H_O\colon \sigma_\pi^2 = 0$$

The corresponding statistic is defined as:

$$F_{Subject(Group)} = \frac{\frac{SS_{S(G)}}{N-s}}{\frac{SS_R}{(N-s)(n-1)}} \sim F_{N-s,(N-s)(n-1)}$$

where $SS_{S(G)}$ and SS_R have been defined above.

Commonly used contrasts for time – The time contrasts discussed in the single-group model are also of interest in the multiple-group model. Orthogonal polynomial contrasts will be discussed here.

Orthogonal Polynomial Partition of SS – Let us define a set of $n-1$ contrasts with $\boldsymbol{c}_j$ representing the 1 by n vector of contrasts of order j (linear, quadratic, …) and in which $\bar{\boldsymbol{y}}_{..}$ is the n by 1 vector of means at each time point (over groups and subjects).

The F-statistic corresponding to the linear trend can be expressed as:

$$F_{T_1} = \frac{SS_{T_1}}{MS_R} \sim F_{1,(N-s)(n-1)}$$

where $SS_{T_1} = N\boldsymbol{c}_1\, \bar{\boldsymbol{y}}_{..}\bar{\boldsymbol{y}}'_{..}\boldsymbol{c}'_1$, $MS_R = \frac{SS_R}{(N-s)(n-1)}$ and, as already discussed in the single-sample model, for a design with four time points $\boldsymbol{c}_1 = [-3 \quad -1 \quad 1 \quad 3]\frac{1}{\sqrt{20}}$.

Note that this expression in terms of vectors is equivalent to that seen in the previous model where a contrast $L_{j'}$ was denoted as a linear combination of coefficients and time point averages $L_{j'} = \sum_{j=1}^{T} c_{j'j}\bar{y}_{.j}$.

Likewise, the statistic to test the quadratic trend can be expressed as:

$$F_{T_2} = \sim F_{1,(N-s)(n-1)}$$

where $SS_{T_2} = N\boldsymbol{c}_2\, \bar{\boldsymbol{y}}_{..}\bar{\boldsymbol{y}}'_{..}\boldsymbol{c}'_2$ and, for a four time point design, $\boldsymbol{c}_2 = [1 \quad -1 \quad -1 \quad 1]\frac{1}{\sqrt{4}}$.

The contrast to test a trend of order $n-1$ can then be generalized as:

$$F_{T_{n-1}} = \sim F_{1,(N-s)(n-1)}$$

where $SS_{T_{n-1}} = N\boldsymbol{c}_{n-1}\,\overline{\boldsymbol{y}}_{..}\overline{\boldsymbol{y}}'_{..}\boldsymbol{c}'_{\boldsymbol{n-1}}$.

In order to find the polynomial of least degree, one can start by testing the polynomial of the highest degree and work backwards towards the lowest degree or linear trend. The SS for the Group by Time interaction can be decomposed as follows:

- $SS_{GT_1} = \sum_{h=1}^{s} N_h\, \boldsymbol{c}_1\, \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.}\boldsymbol{c}'_{\boldsymbol{1}} - SS_{T_1}$ (linear trend),
- $SS_{GT_2} = \sum_{h=1}^{s} N_h\, \boldsymbol{c}_2\, \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.}\boldsymbol{c}'_{\boldsymbol{2}} - SS_{T_2}$ (quadratic trend),
- $SS_{GT_{n-1}} = \sum_{h=1}^{s} N_h\, \boldsymbol{c}_{n-1}\, \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.}\boldsymbol{c}'_{\boldsymbol{n-1}} - SS_{T_{n-1}}$ ($(n-1)th$ trend).

Note that now the degrees of freedom corresponding to these sum of squares is $(s-1)$. The corresponding F-statistics are given by,

$$F_{GT_{n-1}} = \frac{\frac{SS_{T_{n-1}}}{s-1}}{MS_R} \sim F_{s-1,(N-s)(n-1)}, \cdots, F_{GT_1} = \frac{\frac{SS_{T_1}}{s-1}}{MS_R} \sim F_{s-1,(N-s)(n-1)}$$

3.1.6.1.3. Method: One-Sample MANOVA

Before presenting the methodological details for one-sample MANOVA, the data arrangement in the ANOVA versus MANOVA framework will be reviewed so readers can familiarize themselves with the data formatting and indexing of the following sections.

As discussed in the previous models, the main advantage of using MANOVA for longitudinal data is that it assumes a general form for the covariance pattern for the repeated measurements. On the other hand, the main disadvantage is that it requires complete data, i.e. data for all the repeated occasions on which subjects are measured.

The main distinction between ANOVA and MANOVA pertains to the format of the data. In the ANOVA model, each subject-occasion represents one row or observation in the data set. For instance, in a model with $n = 3$ repeated measurements each subject occupies 3 rows in the data set. There may be more variables in the data set, but the purpose of Table 5 is to illustrate that the repeated measures are arranged under one dependent variable y (also called data in *long* format).

Subject	**Time**	$\boldsymbol{y}$
1	1	y_{11}
1	2	y_{12}
1	3	y_{13}
2	1	y_{21}
2	2	y_{22}
2	3	y_{23}
⋮	⋮	⋮

Table 5. *Data structure in long format under ANOVA framework*

The data structure under the MANOVA model differs from the ANOVA in that each subject is represented by only one observation (or row) in the dataset (also called data in *wide* format). This is achieved by representing the dependent variable using n different variables (or columns). Using the same example, Table 6 represents the case of 3 repeated measures in wide format.

Subject	$\boldsymbol{y1}$	$\boldsymbol{y2}$	$\boldsymbol{y3}$
1	$y1_1$	$y2_1$	$y3_1$
2	$y1_2$	$y2_2$	$y3_2$
⋮	⋮	⋮	⋮

Table 6. *Data structure in wide format under MANOVA framework*

It is easy to observe that time is not a variable, but the number of repeated measures is implicit in the number of dependent variables. In fact, the repeated measures are treated as a data vector, hence the multivariate nature of MANOVA. Having covered the MANOVA data format, the one-sample MANOVA is discussed below.

Model assumptions – Let $\boldsymbol{y}_i$ be a n by 1 vector representing the n repeated measures of the response variable. The one-sample MANOVA can be presented as:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i$$

where,

- $\boldsymbol{\mu}$ is the n by 1 vector representing the mean for each time point or repeated measure,
- $\boldsymbol{\varepsilon}_i$ represents the n by 1 vector of errors, distributed as $N(0, \boldsymbol{\Sigma})$.
- The variance-covariance matrix $\boldsymbol{\Sigma}$ of the error term can be of a general form. In other words, there is no such assumption as the CS seen in the univariate case.

Note that in the univariate case, the $\boldsymbol{\Sigma}$ matrix can be expressed as:

$$\boldsymbol{\Sigma} = \sigma_\pi^2 \mathbf{1}_n \mathbf{1}_n' + \sigma_e^2 \boldsymbol{I}_n \text{ ,}$$

where, $\mathbf{1}_n \mathbf{1}_n'$ represents the n by n matrix of ones and $\boldsymbol{I}_n$ is the n by n identity matrix. Likewise, the mean vector can be expressed as $\boldsymbol{\mu} = \mu + \boldsymbol{\tau}$,where μ denotes the grand mean and $\boldsymbol{\tau}$ represents the time effects vector. Therefore, all the ANOVA results can be pulled out from the MANOVA model.

Growth curve analysis – Growth curve analysis, also called polynomial representation, consists of modeling the mean vector as a polynomial function of time:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \beta_1 + + \begin{bmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_n^2 \end{bmatrix} \beta_2 + \cdots + \begin{bmatrix} t_1^{q-1} \\ t_2^{q-1} \\ \vdots \\ t_n^{q-1} \end{bmatrix} \beta_{q-1}$$

where t_1, t_2, …, t_n represents time point values and $q \leq n$ indicates the degree of the polynomial. The model equation can therefore be written using matrix notation as:

$$\boldsymbol{y}_i = \boldsymbol{T}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

It is recommended to orthogonalize $\boldsymbol{T}$ by expressing the mean vector as $\boldsymbol{\mu} = \boldsymbol{P}'\boldsymbol{\theta}$ where $\boldsymbol{P}$ is the q by n matrix of orthogonal polynomials with the first row representing the constant term, the second row, the linear, the third, the quadratic, and so on. This is obtained through the use of the Cholesky decomposition which yields a q by q lower triangular matrix $\boldsymbol{S}$ such that $\boldsymbol{P} = \boldsymbol{S}^{-1}\boldsymbol{T}$ and

$\boldsymbol{SS}' = \boldsymbol{TT}'$. See Pearson and Hartley (1976) for orthogonal polynomial contrasts using equal time intervals.

Some statistical packages such as SAS have procedures that give matrix T into an orthogonal polynomial matrix. For instance, if one were to use this procedure with the following time matrix for $n = 4$,

$$T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 9 \\ 0 & 1 & 8 & 27 \end{bmatrix}$$

The corresponding orthogonal polynomial matrix would be:

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{matrix} \div\sqrt{4} \\ \div\sqrt{20} \\ \div\sqrt{4} \\ \div\sqrt{4} \end{matrix}$$

Note that the elements of each row are divided by the sum of the squares of the row elements, indicated by the division sign on the right of the matrix.

Each row, after being divided by its corresponding square root value, represents the coefficients corresponding to the polynomial contrasts that were presented in the previous models. In this case, the first row represents the constant term, the second row, the linear term, and so on.

The orthogonal polynomial trend model can be written as:

$$\boldsymbol{Py}_i = \boldsymbol{P\mu} + \boldsymbol{P\varepsilon}_i = \boldsymbol{\theta} + \boldsymbol{\varepsilon}_{\boldsymbol{i}}^*$$

where,

- $\boldsymbol{\theta}$ is the n by 1 vector of transformed population means, estimated by the transformed sample means vector $\hat{\theta} = \boldsymbol{P\bar{y}}_.$,
  - $\boldsymbol{\bar{y}}_.$ indicates the n by 1 vector of time point means,
- $\boldsymbol{\varepsilon}_{\boldsymbol{i}}^* \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}^*)$ represents the transformed vector of residuals where $\boldsymbol{\Sigma}^* = \boldsymbol{P\Sigma P}'$.

Notice that the test of sphericity seen in the ANOVA models is equivalent to testing whether the lower $(n-1) \times (n-1)$ partition of the $n \times n$ matrix $\boldsymbol{P\Sigma P'}$ has constant diagonal elements and zero off-diagonal elements.

Moreover, the MANOVA table is usually presented by the following three elements:

1. Sum of Squares for Time:

$$\mathbf{SST}^* = N\boldsymbol{P}\bar{\boldsymbol{y}}_{.}\bar{\boldsymbol{y}}_{.}'\boldsymbol{P}'$$

SST^* represents the sum of squares and cross-product matrix with dimensions n by n. The first element of its diagonal equals $Nn\bar{y}_{..}^2$, and is a function of the grand mean. The other $n - 1$ elements of the diagonal of SST^* correspond to the orthogonal polynomial partition of Time into SS_{T_1} for the linear trend, SS_{T_2} for the quadratic term, and so on. (Note that this has already been explained in the ANOVA case).

2. Residual Sum of Squares:

$$\mathbf{SSR}^* = \boldsymbol{P}(\boldsymbol{Y}'\boldsymbol{Y} - N\bar{\boldsymbol{y}}_{.}\bar{\boldsymbol{y}}_{.}')\boldsymbol{P}'$$

$\boldsymbol{Y}$ is the N by n matrix that contains all data. The first diagonal element of the n by n, SSR^* matrix corresponds to the subjects SS and the other $n-1$ diagonal elements correspond to the orthogonal polynomial decomposition of Error or Subject by Time SS.

3. Total Sum of Square:

$$\mathbf{SSY}^* = \boldsymbol{PY'YP'}$$

If the sphericity assumption is met, it is possible to extract the univariate repeated measures ANOVA results from the SST^* and SSR^* matrices (for further details refer to Hedeker and Gibbons, 2006).

Multivariate test of the time effect – Testing the null hypothesis of no time effect is equivalent to testing whether all elements of the n by 1 mean vector $\boldsymbol{\mu}$ are equal, or whether the n by 1 vector of time effects equals a vector of zeros $H_0\colon \boldsymbol{\tau} = \mathbf{0}.$ In order to test this hypothesis, the elements of the lower $(n-1) \times (n-1)$ submatrices of SST^* and SSR^* need to be extracted. Under the null hypothesis, both submatrices have the same expectation. Therefore, the same logic used in the univariate F-test (where the mean squares for time and residual are compared) is used here with the corresponding SS matrices. The equivalent to a ratio of MS is solving the following determinant:

$$\left|\mathrm{SST}^*_{(n-1)} - \lambda\, \mathrm{SSR}^*_{(n-1)}\right| = 0$$

which has a nonzero eigenvalue (or latent root) λ_1. This eigenvalue will be equal to one if H_0 is "true" (i.e. when $\mathrm{SST}^*_{(n-1)} = \mathrm{SSR}^*_{(n-1)}$).

Solving this equation yields a series of overall tests statistics such as Wilk's Lambda, the Hotelling-Lawley trace, and the Pillai-Barlett trace. Under the null hypothesis, all these test statistics approximately follow an F-distribution.

Test of Specific Time Elements – If sphericity holds, then one can use the univariate RM ANOVA tests, whose numerators can be extracted from the lower n – 1 diagonal elements of SST^*. The MR_R would be used as common denominator for all trend contrasts:

$$F_1 = \frac{SST_1}{\frac{SS_R}{(N-1)(n-1)}}, F_2 = \frac{SST_2}{\frac{SS_R}{(N-1)(n-1)}}, \dots, F_{n-1} = \frac{SST_{n-1}}{\frac{SS_R}{(N-1)(n-1)}}$$

If sphericity is not met and the MANOVA is conducted, the test of the specific trend components are built using their corresponding error term extracted from the SSR^* submatrix.

$$F_1 = \frac{SST_1}{\frac{SSR_1}{N-1}}, F_2 = \frac{SST_2}{\frac{SSR_2}{N-1}}, \dots, F_{n-1} = \frac{SST_{n-1}}{\frac{SSR_{n-1}}{N-1}}$$

Note that now each denominator has only $N-1$ degrees of freedom. This is the reason why, if sphericity holds, the ANOVA tests are more powerful compared to those corresponding to the MANOVA.

3.1.6.1.4. Method: Multiple Samples MANOVA

Model assumptions – Let us assume that there exist:

- $h = 1, \cdots, s$ groups,
- $i = 1, \dots, N_h$ subjects in group h,
- $j = 1, \cdots, n$ time points.

The total number of subjects is defined by $N = \sum_{h=1}^{s} N_h$. Notice that the number of subjects per group N_h can vary, but the number of time points a subject is measured, n, is the same across subjects.

This model can then be expressed as:

$$\boldsymbol{y}_{hi} = \boldsymbol{\mu} + \boldsymbol{\gamma}_h + \boldsymbol{\varepsilon}_{hi}$$

where,

- $\boldsymbol{\mu}$ represents the n by 1 vector of time point means,
- $\boldsymbol{\gamma}_h$ represents the n by 1 vector of group representing the effect of group h,
- $\boldsymbol{\varepsilon}_{hi}$ represents the n by 1 vector of errors with distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ for each of the populations (i.e. the population from which each group h of subjects is drawn).

One important assumption of the multiple groups MANOVA is the homogeneity of variance-covariance assumptions. This means that the same general $\boldsymbol{\Sigma}$ is assumed for all groups. Applying the orthogonal transformation for time, the model can be rewritten as:

$$\boldsymbol{P}\boldsymbol{y}_{hi} = \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{P}\boldsymbol{\gamma}_h + \boldsymbol{P}\boldsymbol{\varepsilon}_{hi}$$

$$\boldsymbol{\varepsilon}_{hi} \sim N(\boldsymbol{0}, \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}')$$

As seen in the one-sample MANOVA, the following step would be to test the transformed $\boldsymbol{\Sigma}^* = \boldsymbol{P\Sigma P'}$ for sphericity. If sphericity is satisfied then the univariate tests are recommended; otherwise, the MANOVA tests should be used.

The MANOVA formulation is given by:

a. Sum of Squares for Time:

$$\mathbf{SST}^* = N\boldsymbol{P}\overline{\boldsymbol{y}}_{..}\overline{\boldsymbol{y}}'_{..}\boldsymbol{P}'$$

where the n by n matrix $\mathbf{SST}^*$is a function of the cross-product matrix from the overall mean vector of repeated measures $\overline{\boldsymbol{y}}_{..}\overline{\boldsymbol{y}}'_{..}$

b. Sum of Squares for Group:

$$\mathbf{SSG}^* = \boldsymbol{P}(\sum_{h=1}^{s} N_h\ \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.} - \mathbf{SST})\boldsymbol{P}'\ = \boldsymbol{P}(\sum_{h=1}^{s} N_h\ \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.} - \overline{\boldsymbol{y}}_{..}\overline{\boldsymbol{y}}'_{..})\boldsymbol{P}'$$

where n by n matrix $\mathbf{SSG}^*$ is a function of the sum of cross-product matrices from the group mean vectors of repeated measures $\sum_{h=1}^{s} N_h\ \overline{\boldsymbol{y}}_{\boldsymbol{h}.}\overline{\boldsymbol{y}}'_{\boldsymbol{h}.}$

c. Residual Sum of Squares:

$$\mathbf{SSR}^* = \boldsymbol{P}(\boldsymbol{SSY} - \boldsymbol{SSG} - \boldsymbol{SST})\boldsymbol{P}'$$

d. Total Sum of Square:

$$\mathbf{SSY}^* = \boldsymbol{PY'YP'}\text{=}\boldsymbol{P}\mathrm{SSY}\boldsymbol{P}' = \boldsymbol{P}(\textstyle\sum_{\boldsymbol{h}}\sum_{\boldsymbol{i}} \boldsymbol{y_{hi}y'_{hi}})\boldsymbol{P}'$$

Following the orthogonal polynomial parameterization, the statistics in the cross-product matrices can be written as:

Time ($df\ = 1$):

$$\mathbf{SST}^* = \begin{array}{c} \begin{bmatrix} SST_0 & & & & \\ & SST_1 & & \cdots & \\ & & SST_2 & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & SST_{n-1} \end{bmatrix} \end{array} \begin{array}{c} \text{constant} \\ \text{linear} \\ \text{quatric} \\ \vdots \\ (n-1)\text{th time} \end{array}$$

Between groups ($df\ = s - 1$):

$$\mathbf{SSG}^* = \begin{array}{c} \begin{bmatrix} SSG_0 & & & & \\ & SSG_1 & & \cdots & \\ & & SSG_2 & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & SSG_{n-1} \end{bmatrix} \end{array} \begin{array}{c} \text{groups} \\ \text{groups} \times \text{linear} \\ \text{groups} \times \text{quatric} \\ \vdots \\ \text{groups} \times (n-1)\text{th time} \end{array}$$

Subjects within groups ($df = N - s$):

$$\mathbf{SSR}^* = \begin{bmatrix} SSR_0 & & & & \\ & SSR_1 & & \cdots & \\ & & SSR_2 & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & SSR_{n-1} \end{bmatrix} \begin{array}{c} \text{subjects in groups} \\ \text{subjects in groups} \times \text{linear} \\ \text{subjects in groups} \times \text{quatric} \\ \vdots \\ \text{subjects in groups} \times (n-1)\text{th time} \end{array}$$

Note that each of the lower $n-1$ diagonal elements in $\mathbf{SST}^*$ corresponds to the orthogonal SS partition of Time. The first diagonal element in $\mathbf{SSG}^*$is used to test for group effect and the lower $n-1$ diagonal elements are utilized for testing the interaction effect between groups and each of the time trends.

All three matrices SST^*, SSG^*, and SSR^* are symmetric. As in the one-sample case, if sphericity is met, the univariate repeated measures results can be pulled out from these matrices (see Hedeker & Gibbons, 2006, for further details).

Multivariate tests – There are two multivariate tests in this MANOVA model, the first being the group by time interaction test. This is achieved by pulling out the lower $(n-1) \times (n-1)$ submatrices of SSG^* and SSR^* and solving the following matrix expression:

$$\left|\mathrm{SSG}^*_{(n-1)} - \lambda\, \mathrm{SSR}^*_{(n-1)}\right| = 0$$

Common statistics provided by software are Wilk’s Lambda, Hotelling-Lawley Trace, and Pillai’s Trace. A non-significant overall group by time test means that the overall test of time effect will be conducted in the same fashion that was seen in the one-sample MANOVA.

Test of Specific Group by Time and Time Components – In contrast to the one-sample case, where the pooled MS_R was used as the denominator, in this model testing for time effects and group by time effects involves using separate denominators.

Following the multivariate Time by Group test, the individual components are tested using the following statistics:

$$F_{GT2} = \frac{\frac{SSG_2}{s-1}}{\frac{SSR_2}{N-s}} \text{ (group by linear trend)}$$

$$F_{GT2} = \frac{\frac{SSG_2}{s-1}}{\frac{SSR_2}{N-s}} \text{ (group by quadratic trend)}$$

$$\cdots$$

Notice that each of above tests follows an $F(s-, N-s)$ under the null hypothesis.

3.2. Advanced Analysis Approaches for Longitudinal Data

In this section, the more advanced methods for analyzing longitudinal data are discussed. The first three models reviewed are the families for the analysis of continuous and discrete repeated measure data using the extensions of generalized linear models (Diggle et al., 2002; Verbeke & Molenberghs, 2009; Fitzmaurice et al., 2004; Molenberghs & Verbeke, 2005). The three model families are:

1. Marginal models: Outcomes are modeled marginalized over all other variables.
2. Subject-specific models: Outcomes are assumed independent, given a collection of subject-specific parameters. The subject-specific methods that will be discussed here are the mixed-effects models (including heterogeneity models and generalized LMMs).

3. Conditional models: Any outcomes within the sequence of measurement occasions are modeled conditionally on other past outcomes. The specific conditional model that will be discussed here is the transition models.

The final models discussed in this section will be advanced analytical techniques for longitudinal data, such as autoregressive models and latent growth curve modeling (including latent class growth models and latent growth mixture models) within the SEM framework, time-series analysis, non-linear and non-parametric modeling.

3.2.1. Review: Marginal Models via Generalized Estimating Equations

Marginal approaches are used when the focus of a study is to examine the effects of variables on the population mean (Edwards, 2000). The marginal model analyzes the relationship between the outcome and the predictors without accounting for between-subject heterogeneity, and the coefficients of the marginal models have a population-level interpretation (rather than an individual-level interpretation); the model is therefore also referred to as the population-average model (Molenberghs & Verbeke, 2005). The term "marginal" indicates that the mean response of the marginal model depends solely on the variables of interest, but not on any random effects and/or past outcomes (Fitzmaurice et al., 2004). This is in contrast to mixed-effects models (discussed in Section 3.2.2.), where the mean response of the model depends on both the variables of interest and the random effects (Fitzmaurice et al., 2004).

Marginal models do not impose distributional assumptions, which is advantageous as very often the outcome variables may be discrete and the usual normality assumption would be hard to attain (Fitzmaurice et al., 2004). A variety of marginal models exist; however, they are computationally expensive due to high dimensional vectors of correlated data making parameter estimation via the maximum likelihood undesirable (Fitzmaurice et al., 2004). As a consequence, an alternative estimating method, the GEEs, was proposed by Liang and Zeger (1986). Within the GEE framework, the dependency correction of observations is done by implementing a certain "working" covariance pattern for the repeated measurements of the outcome (Liang & Zeger, 1986). Note that, even when the working covariance structure is incorrect, the GEE method would still yield unbiased parameter estimates (Liang & Zeger, 1986). In sum, the two main advantages of the GEE modeling are its robustness to misspecification of the repeated measures' covariance pattern and the simplicity of its computations (Fitzmaurice et al., 2004; Molenberghs & Verbeke, 2005).

Estimating marginal models for longitudinal data via GEE has been widely used in Education and Psychology. For example, Cardozo et al. (2012) fitted GEE longitudinal models to study whether there were any associations between the outcome variables (i.e., anxiety, depression, burnout, emotional exhaustion, burnout depersonalization, burnout personal accomplishment, and life satisfaction) and predictive factors of interest (e.g., gender, age, marital status, job function, hardship assignment, mental illness history, trauma exposure, social support, motivation, child trauma, extraordinary stress, health habits index, and adult trauma) over the study period (i.e., pre-deployment, post-deployment, and 3–6 months after deployment) for international humanitarian assistance employees providing care in crises. Other examples of the GEE applications for longitudinal data in the field of Education and Psychology included Kent et al. (2011); Van Nguyen, Laohasiriwong, Saengsuwan, Thinkhamrop and Wright (2015); Lee et al. (2016); Boden, Van Stockum, Horwood and Fergusson (2016); and Moskowitz et al. (2017).

3.2.1.1. Method: Generalized Estimating Equations

Generalized Linear Models (GLMs) – Since GEE is an extension of GLM for correlated data (e.g. longitudinal data), it is necessary to revisit the Generalized Linear Models (GLMs) before introducing the GEE models. GLMs develop a family of models, under which various regression methods can be defined as special cases. These different forms of regression models can vary with regard to their outcome variable, which is assumed to originate from a class of distributions called the exponential family. Therefore, this unitary framework includes linear regressions (continuous response variable), Logistic regressions (binary response variable), and Poisson or negative binomial regressions (count response variable).

Model specifications - A GLM is defined by:

1. A *linear predictor,* $\eta_i = x_i'\beta$. In other words, a linear combination of covariates, x_i', and regression coefficients β for subject i.
2. A *link function* $g(.)$. This translates the expected value of the dependent variable, $\mu_i = E(y_i)$, into the linear predictor $g(\mu_i) = \eta_i$. For instance, in the linear multiple regression, the link function happens to be the identity link since $g(\mu_i) = \mu_i = \eta_i$. In other words, for linear regression, the expected value of the outcome variable is a linear combination of the predictors. The Logistic regression, used when the dependent variable is a binary outcome, can be written as:

$$log\left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = x_i'\beta$$

Note that since $P(y_i = 1) = E(y_i) = \mu_i$, the Logistic regression uses the following link function $g(\mu_i) = \log(\frac{\mu_i}{1-\mu_i})$. Poisson regression is used when the outcome variable is count and is written as μ_i=exp($x_i'\beta$). The link function is $g(\mu_i) = \log \mu_i$.

3. The *form of the conditional variance* of *outcome* given the predictors $V(y_i) = \phi v(\mu_i)$, where,
   a. $v(\mu_i)$ denotes a variance function which is known,
   b. ϕ denotes the scale parameter, which can either be known or estimated.

For instance, for the linear regression model $\upsilon(\mu_i) = 1$ and ϕ represents the error variance. For the Logistic regression, $\upsilon(\mu_i) = \mu_i(1 - \mu_i)$ and ϕ is set to 1. In the case of the Poisson distribution, where the mean and variance are equal, $\upsilon(\mu_i) = \mu_i$ and ϕ is again set to 1. An exception is for methods that account for under- or over-dispersion, such as the negative binomial regression. For these models, ϕ is estimated.

The GEE Models – An important characteristic of the GEE models is that only the marginal distribution of y at each time point needs to be specified. Therefore, it avoids the need of using multivariate distributions. A very attractive aspect of the GEE models is that, even when the covariance structure of the repeated measures is mis-specified, they produce consistent and asymptotically normal estimates of the regression coefficients.

Model specifications for the GEE model – As with the GLMs, first the linear predictor is specified as a linear combination of the variables, $\eta_{ij} = x'_{ij}\beta$, where x'_{ij} indicates the vector of variables for subject i at time j. It follows a link function, $g(\mu_{ij}) = \eta_{ij}$, which will depend on the type of response (continuous, binary, or count). A third specification shared with the GLMs is that the variance is defined as a function of the mean $V(y_{ij}) = \phi\upsilon(\mu_{ij})$.

An additional specification of the GEE models is the *working correlation structure* for the repeated measures which is an n by n correlation matrix, where n is the number of time points. Subjects do not need to be measured at all time points, therefore each subject will have his/her own correlation matrix $\boldsymbol{R}_i$ of size $n_i \times n_i$ with $n_i \leq n$. The individual correlation matrix $\boldsymbol{R}_i$ is written as a function of a vector of parameters $\boldsymbol{a}$. Although GEE is robust to misspecifications of the covariance pattern, it is recommended to choose an $\boldsymbol{R}$ consistent with the observed correlations. If the choice of $\boldsymbol{R}$ is incorrect, the estimators are less efficient.

Common working correlation forms can be listed as follows:

1. *Independence:* The simplest form is represented by an n by n identity matrix,

$$\boldsymbol{R}_i(a) = \boldsymbol{I}$$

As the name indicates, the independence form assumes that the repeated measures are not correlated. This assumption is not realistic for longitudinal data.

2. *Exchangeable:* This form is the second simplest and it assumes that all correlations are equal across time,

$$\boldsymbol{R}_i(a) = \rho$$

This assumption is equivalent to the compound symmetry (CS) for Covariance Pattern Modeling (covered in section 3.2.8.1.).

3. $AR(1)$*:* The first-order autoregressive form is less restrictive than the previous one because it assumes that the correlation between repeated measures is an exponential function of the lag,

$$\boldsymbol{R}_i(a) = \rho^{|j-j'|} \quad \text{for} \quad j \neq j'$$

4. *Toeplitz or m-dependent:* This structure assumes that all correlations within a time lag are equal, but, in contrast to $AR(1)$, here lags of different orders have no functional relationships between them. This form is written as,

$$\boldsymbol{R}_i(a) = \rho_{|j-j'|} \quad \text{if} \quad j - j' < m$$

$$\boldsymbol{R}_i(a) = 0 \qquad \text{if} \quad j - j' > m$$

where the fullest structure in which all lagged correlations are estimated is when $m = n - 1$. Notice that this structure is less restrictive than the $AR(1)$ for which only one term, ρ, is estimated.

5. *Unstructured:* Under this structure, $\frac{n(n-1)}{2}$ parameters are estimated. This structure is the most efficient, but only useful when there are few time points. For large n and under the presence of missing data, the estimation of $\boldsymbol{R}$ can become quite complicated.

GEE Estimation – Let us define $\boldsymbol{A}_i$ as an n by n diagonal matrix with diagonal elements $V(\mu_{ij})$ and $\boldsymbol{R}_i(a)$ as the working correlation matrix. Then the working variance-covariance matrix for subject i is expressed as:

$$V(a) = \phi \boldsymbol{A}_{\boldsymbol{i}}^{\frac{1}{2}} \boldsymbol{R}_i(a) \boldsymbol{A}_{\boldsymbol{i}}^{\frac{1}{2}}$$

The GEE estimator of β is attained by solving the following equation:

$$\sum_{i=1}^{N} \boldsymbol{D}_i' [V(\hat{a})]^{-1} (y_i - \mu_i) = \mathbf{0}$$

where $\hat{a}$ is a consistent estimate of a and $D_i = \frac{\partial \mu_i}{\partial \beta}$.

Solving GEE, which is done as an iterative process, involves repeating the following steps until convergence is achieved:

1. Compute estimates of β given the estimates of $\boldsymbol{R}_i(a)$ and ϕ using iteratively reweighted least squares (IRLS).
2. Based on the obtained estimates of β, compute estimates of a and ϕ. This is achieved by calculating the standardized residuals,

$$r_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\sqrt{[V(\hat{a})]_{jj}}}$$

Once convergence is achieved, the standard errors associated with the estimated β are of interest to conduct hypothesis testing. Two versions of standard errors can be computed for GEE models (see Hedeker & Gibbons, 2006 for further details).

3.2.2. Review: Mixed-Effects Models

A more advanced but general treatment of repeated measure data requires more rigorous approaches; these methods have been developed by researchers over the past 30 to 40 years. The most commonly used method is the mixed-effects regression model (Laird & Ware, 1982). Mixed-effects modeling is essentially regression analysis that allows two kinds of effects: (a) fixed effects, which can be used to describe the population, and (b) random effects, which can be used to capture correlations of repeated measures and describe the variability across subgroups of the sample and/or the cluster-specific trends over time (Fitzmaurice et al., 2004). Mixed-effects models are subject-specific methods, which are differentiated from marginal models (or population-averaged models) by the inclusion of subject-specific parameters (Molenberghs & Verbeke, 2005). Subject-specific approaches are most beneficial when the focus of the research is to make inferences about individuals rather than the population average (Diggle et al., 2002). The premise of a mixed-effects model (for both Gaussian continuous responses or discrete/non-Gaussian responses) is that there is a naturally occurring heterogeneity across individuals, which can be represented by a probability distribution (Diggle et al., 2002).

The versatility of mixed-effects modeling has led to a variety of terms for the models it makes possible in different disciplines. Because of the simultaneous development of mixed-effects models across many fields, the models have been known under many different names, including random coefficient models, random-effects models, random intercept models, random regression models, mixed-effects models, multilevel models, hierarchical linear models (HLMs), and variance component models (Holden, Kelley, & Agarwal, 2008; Gibbons et al., 2010; Garson, 2013; Woltman, Feldstain, MacKay, & Rocchi, 2012; Ker, 2014; Lininger, Spybrook, & Cheatham, 2015). In spite of many different labels, the commonality of these methods is the inclusion of random-subject effects into the regression to account for subject-specific effect. This allows for the description of an individual's trend across time, explains the degree of individual variation that exists in the population of individuals, and yields the correlational structure of the repeated measure data (Gibbons et. al, 2010; Garson, 2012). It should be also noted that in a linear mixed-effects model, it is assumed that the conditional distribution of each observation, given a vector of random effects, has a normal distribution (Fitzmaurice et al., 2004). In addition, the random-effects of the mixed-effects models are assumed to have a multivariate normal distribution (Fitzmaurice et al., 2004). Examples of mixed-effects models in Education and Psychology include Sitzmann and Ely (2010); Brown et al. (2012); Shephard et al. (2015); and Sullivan, Kohli, Farnsworth, Sadeh and Jones (2017).

The primary advantages of mixed-effects models include:

1. The ability to include both time-invariant predictors such as country of birth and time-varying predictors such as age in the modeling process;
2. Participants are not expected to be observed on the same number of time points, and hence individuals with missing data across repeated measures are included in the analysis (that is, irregularly timed and missing data can be handled by the models without the need for explicit imputation);
3. Such models allow multilevel hierarchical modeling which enables predictions at each hierarchy level (Gibbons et al., 2010; Woltman et al., 2012; Lininger et al., 2015).

Hierarchical data or multilevel data are a commonly occurring phenomenon in Educational and Psychological research (Woltman et al., 2012; Ker, 2014). Hierarchical data means that measurements at lower levels are nested within higher level units (Ker, 2014). For example, in the education sector, data can be collected and organized at student, classroom, school, and district levels. Each subject's observations collected over time are nested within the individual is another type of hierarchical data; the repeated measures are nested within each person (Ker, 2014).

Mixed-effects models for multilevel analysis address hierarchically nested data structures, which often are termed HLM. These models account for the fact that subjects within a specific group may be more similar than subjects in other groups (Garson, 2013). Additionally, these models can investigate both lower- and higher-level unit variance corresponding to the outcome (Ker, 2014). In sum, HLMs allow the researchers to explore the associations *within* a certain hierarchical level, as well as associations *between* (or *across*) hierarchical levels, at the same time (Woltman et al., 2012; Ker, 2014). HLMs/Mixed effects models are essential tools for analyzing hierarchically structured data in Psychological and Educational research (Ker, 2014). Jang, Reeve, and Deci (2010) collected hierarchically structured data for students' individual self-reported engagement, where the self-reported engagement questionnaires were completed by 1584 students in 84 classrooms within nine schools. To analyze these data, "on the first level (between students' level), using HLM, regression equations were modeled to detect engagement differences among students sitting in the same classroom. At the second level (between-teachers level), regression equations were modeled for characteristics that differed between teachers (autonomy support, structure). At the third level (between-schools level), regression equations were modeled for the different schools in which the teachers taught" (Jang, Reeve, & Deci, 2010). Other examples of mixed-effects models for multilevel analysis addressing hierarchical data in Education and Psychology can be found in Han, Capraro and Capraro (2015); Baker, Tichovolsky, Kupersmidt, Voegler-Lee and Arnold (2015); Kisa and Correnti (2015); and Kwok et al. (2018).

3.2.2.1. Methods: Linear Mixed Models

We have seen that traditional models, such as RM ANOVA and RM MANOVA, have important limitations including restrictive covariance structures (e.g. CS) and the inability to handle missing data. Another common limitation is that individuals are supposed to be measured at the same time points. On the other hand, the Mixed-effects Regression Models (MRMs) overcome these limitations by modeling specific individual trends using what is called a random effect. These types of models are very useful in longitudinal data analysis since subjects with incomplete data can be included in the model.

This section will focus only on models that have a continuous response outcome, also called Linear Mixed Models (LMMs).

A natural way to introduce the LMM is by starting from a simple linear regression model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + e_{ij}$$

where y represents the response variable and t the time variable, which is considered to be continuous. The subscripts i and j indicate that both the response and time variables vary across subjects $i = 1,2 \cdots N$ and occasions $j = 1,2 \cdots n_i$.

Under the linear regression approaches, the error terms are assumed to be distributed independently under $N(0, \sigma^2)$. Given that subjects are measured on more than one occasion, the independence assumption is unreasonable for repeated measure data. This model also assumes the slope for time is the same for all individuals, which is not realistic either. LMMs add a specific-individual effect to account for the clustering in the data (occasions nested within individuals) and allow the estimation of individual time trends.

3.2.2.1.1. Model: Random Intercept

The simplest extension of the regression model consists of adding a specific-subject intercept to the model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varsigma_{0i} + e_{ij},$$

$$e_{ij} | \varsigma_{0i} \sim N(0, \sigma^2)$$

where ς_{0i} represents the deviation of individual i from the population average. It can also be understood as the influence of individual i on his/her own repeated measures.

The formulation above, into one single equation, is known as *reduced form*. The model can also be expressed in a *hierarchical form* or two-stage formulation, where the model is split into a within-subjects or level-1 model,

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + e_{ij}$$

and the between-subject or level-2 model,

$$b_{0i} = \beta_0 + \varsigma_{0i}$$

$$b_{1i} = \beta_1$$

The reduced form indicates that the outcome for subject i at time j is explained by that subject's baseline level b_{0i} and a time trend indicated by the slope b_{1i}. The level-2 model provides an equation for the baseline level for subject i which is determined by a population baseline level β_0 plus the specific subject i contribution ς_{0i}.

Therefore, this model allows a specific initial level for each individual, represented by b_{0i}. However, the model assumes that the slope is the same for all individuals. One way to conceptualize this is to imagine that each individual is represented by a regression line, which is parallel to the population trend, the only difference between the individual trends being determined by ς_{0i}.

Note that if level-1 and level-2 models are combined, one obtains the reduced form equation.

Model assumptions – The subject-specific effects, ς_{0i}, are treated as random effects since the sample of subjects is assumed to be representative of a larger population.

The error term e_{ij} is assumed to follow $N(0, \sigma_{\varsigma}^2)$. Particularly, the error term is *conditionally independent* and $N(0,\ \sigma^2)$. Note that conditional independence means conditional on the random subject-specific effect ς_{0i}.

The error terms ς_{0i} and e_{ij} are sometimes referred to as *permanent* and *transitory* components as ς_{0i} represents the time invariant characteristics of the individuals and e_{ij} the moment's random deviation.

A random-intercept model with assumed independent errors, as in the one presented here, implies a CS pattern for the variance and covariance matrix. In other words, both variance and covariance are assumed constant across time:

$$V(y_{ij}) = \sigma_{\varsigma}^2 + \sigma^2$$

$$C(y_{ij}, y_{ij'}) = \sigma_{\varsigma}^2, \quad \text{for } j \neq j'$$

The intraclass correlation is expressed as the ratio of the random intercept variance σ_{ς}^2 to the total variance $\sigma_{\varsigma}^2 + \sigma^2$. The interpretation of the ICC is the same as the one seen for the RM ANOVA model; it denotes the proportion of variance due to individuals. Note that this model allows the use of less restrictive assumptions for the variance-covariance matrix, such as autoregressive, moving averages structures, or an unstructured form.

Finally, it is worth highlighting again that in LMMs, each individual is measured on n_i occasions, which means that individuals with missing data for some time points are still included in the analysis. Also, the subscript i for the time variable indicates that each individual can be measured at different occasions. In other words, each subject can be measured at his/her own individual schedule (e.g. follow-up visits).

3.2.2.1.2. Model: Random Coefficient

The random intercept model may be not suitable for repeated measure data since it assumes that the rate of change is the same for all subjects and that all measurements of the same subject will have the same degree of correlation, regardless of their proximity in time. An extension of the random intercept model is therefore to allow both intercept and slope (time trend) to vary across subjects. This can be expressed in the level-2 model by adding another error term to the time slope as follows:

$$b_{0i} = \beta_0 + \varsigma_{0i}$$

$$b_{1i} = \beta_1 + \varsigma_{1i}$$

The level-1 equation does not change with respect to the random intercept model, hence

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + e_{ij}$$

This new error component term, ς_{1i}, also known as random coefficient, can be interpreted as the slope deviation for subject i, in the same way that ς_{0i} is considered the intercept deviation for subject i. The reduced form can be obtained by substituting the level-2 model in the level-1 model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varsigma_{0i} + \varsigma_{1i}t_{ij} + e_{ij}$$

Model assumptions - The errors e_{ij} are conditionally independently distributed as $N(0, \sigma^2)$. Their independence is conditional on ς_{0i} and ς_{1i}.

The random intercept and the random coefficient are assumed to follow a bivariate normal distribution with random-effect variance and the following covariance matrix:

$$\Sigma_\varsigma = \begin{bmatrix} \sigma^2_{\varsigma_0} & \sigma_{\varsigma_0\varsigma_1} \\ \sigma_{\varsigma_0\varsigma_1} & \sigma^2_{\varsigma_1} \end{bmatrix}$$

The variance component $\sigma^2_{\varsigma_0}$ specifies the amount of heterogeneity in the individual intercepts or deviation from the population intercept. Likewise, the variance term $\sigma^2_{\varsigma_1}$ indicates the heterogeneity in slopes, i.e. how much individual slopes differed from the population slope represented by β_1. The covariance term $\sigma_{\varsigma_0\varsigma_1}$ indicates how the random intercept and slope vary together. For instance, a positive covariance is suggestive that subjects with greater baseline measurements are expected to have greater positive slopes.

The exact manner in which time is coded is important since it will affect the interpretation of the model coefficients. For example, if time is coded starting with 0 for baseline and followed by unit increments for follow-up measurements (i.e. 0, 1, 2, 3, 4, …), the intercept parameters in the model β_0 and ς_{0i} refer to the baseline population and individual values. However, if time is centered (i.e. the average of time is subtracted from each time value), then the interpretation of the intercept parameters refers to the midpoint.

3.2.2.1.3. Model: Random Coefficient with a Time-Invariant Covariate

In the previous model, time was defined as a continuous and, needless to say, time-variant covariate. Let us assume that one is interested in adding a time-invariant covariate x_i (e.g. gender). Any level-2 covariates (i.e. characteristics that do not vary across time) will be included in the level-2 model. For simplicity, x_i is assumed to be a binary variable. (Note that for categorical variables with k groups with $k > 2$, $k-1$ dummies need to be included in the model.) The level-2 model equation corresponding to the intercept can be written as:

$$b_{0i} = \beta_0 + \beta_2 x_i + \varsigma_{0i}$$

If the time-invariant covariate x_i only affects the individual intercept (e.g. assuming that the baseline salary for women is lower than that of men, but that the trend is the same for both), this would be the only change to the level-2 model. The reduced form for this model can be written as:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \varsigma_{0i} + \varsigma_{1i} t_{ij} + e_{ij}$$

However, it seems more realistic to consider that the individual slope will be also affected by the covariate x_i. So in the prior example, it seems realistic to assume that men's salaries will grow at a greater rate than that of women. In this case, the level-2 second equation will be written as:

$$b_{1i} = \beta_1 + \beta_3 x_i + \varsigma_{1i}$$

Substituting the level-2 equations in the level-1 model leads to what is called *cross-level interaction* in the reduced form:

$$y_{ij} = \underbrace{\beta_0 + \beta_2 x_i + \varsigma_{0i}}_{b_{0i}} + \underbrace{(\beta_1 + \beta_3 x_i + \varsigma_{1i})}_{b_{1i}} t_{ij} + e_{ij}$$

$$= \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 x_i t_{ij}}_{fixed} + \underbrace{\varsigma_{0i} + \varsigma_{1i} t_{ij} + e_{ij}}_{random}$$

3.2.3. Review: Two Extensions of Mixed-Effects Models

Recall that in a mixed-effects approach, it is assumed that (a) the conditional distribution of each observation, given a vector of random effects, has a normal distribution; and (b) the random effects are assumed to have a multivariate normal distribution (Fitzmaurice et al., 2004). Two extensions of mixed-effects models are discussed in this section, including Generalized LMM and Heterogeneity models, to relax these two assumptions.

3.2.3.1. Review: Generalized Linear Mixed Models

Many outcome variables that are of interest in Education and Psychology are nominal variables with two or more categories, such as school achievement, dropout status, or self-reported satisfaction level. The generalized LMM is the most commonly used random effects model in the context of discrete repeated measurement (Molenberghs & Verbeke, 2005). In a generalized LMM, it is assumed that (a) the conditional distribution of each observation, given a vector of random effects that belongs to the class of exponential family; and (b) the random effects are assumed to have some type of multivariate distribution (multivariate normal distribution is commonly assumed in practice) (Fitzmaurice et al., 2004). The generalized LMM is also an extension of a GLM to include both fixed and random effects, thus the distribution of the response is defined conditionally on the random effects (Molenberghs & Verbeke, 2005). Methods for estimating generalized LMMs have appeared in the literature (Hartzel, Agresti, & Caffo, 2001), and some statistical details will be covered shortly. Yet because the procedures for estimations are complex (and beyond the scope of this dissertation), full statistical details of this method will not be provided here. Examples of the utilization of generalized LMMs in the fields of Education and Psychology include Neighbors et al. (2010); Christens and Speer (2011), Piasecki et al. (2014); Monfort, Howe, Nettles and Weihs (2015); and Kinnunen et al. (2016).

3.2.3.1.1. Methods: Generalized Mixed Models (GMM)

GMM is the extension of LMMs to categorical dependent variables. It has become an active area of research, particularly in medical fields where categorical outcomes are very common. This section will review the Logistic regression, a popular choice for binary/dichotomous outcomes, and the Poisson model, which is used for count data.

3.2.3.1.1.1. Method: Logistic Regression Model and Mixed-Effects Logistic Regression

In order to introduce the mixed-effects generalization of the Logistic regression model, the traditional Logistic regression (i.e. the fixed-effects Logistic regression) is reviewed first.

Let p_i indicate the probability of an event of interest ($Y_i = 1$) for subject i. The probability of the event not happening ($Y_i = 0$) is $1 - p_i$.

Let $\boldsymbol{x}_i = (1, x_{i1}, \cdots, x_{ip})'$ denote the $(p+1)$ by 1 vector of predictors for subject i with its corresponding regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$. The Logistic regression model can be presented as:

$$p_i = \Pr(Y_i = 1) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})}$$

Another way to write this model is in terms of the log odds (or Logit of the probabilities) as follows:

$$\log\left[\frac{p_i}{1-p_i}\right] = \boldsymbol{x}_i'\boldsymbol{\beta}$$

The advantage of the Logit function, also known as the link function, is that the log odds is linear in its relationship with the explanatory variables and, as such, it shares some of the attractive features of a linear regression model.

Because the model is linear in terms of the Logit, the interpretation of the model coefficients is also expressed in terms of Logits. Therefore, the intercept β_0 is understood as the log odds of the event of interest for a subject with all covariates set to zero ($\boldsymbol{x}_i = \boldsymbol{0}$), and the coefficient β_p represents the change in the log odds for a unit change in the explanatory variable x_p, holding all other variables constant. More commonly, the coefficients are expressed as odds ratios by using the exponential transformation $\exp(\beta_p)$. The transformed coefficient $\exp(\beta_p)$ is then interpreted as the ratio of the odds of the event $Y = 1$ for a unit change in x_p.

Latent variable model – Binary response regression models can also be expressed using the *threshold concept*, which assumes that a continuous latent variable *y* lies beneath the measured dichotomous outcome Y. The values of Y are then determined by a threshold c in the following way: Y equals 1 if $y > c$ and Y equals 0 if $y \leq c$.

The binary response regression model in terms of the latent variable y can be expressed as:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i$$

Note that in this formulation, the error term is included. This error term follows a standard Logistic distribution with mean equal to 0 and a variance of $\frac{\pi^2}{3}$ in the case of the Logistic regression, and $N(0, 1)$ in the case of the Probit regression (the Probit regression is an alternative to model binary response, which is less common than the Logistic regression). In particular, given the different distributions of the error terms, the regression coefficients attained from the Logistic regression are approximately 1.7 times those obtained from the Probit regression (J. S. Long, 1997).

Finally, it is important to highlight that although the formulation above seems identical to that of a regression model for a continuous outcome, the error variance in the latent variable model is fixed, not estimated.

Having provided the necessary background for the Logistic regression, the Mixed-Effects Logistic Regression will now be covered. One of the assumptions of the fixed-effect Logistic regression is that the observations are independent. As mentioned before, this is not the case in longitudinal data, where repeated measures are obtained from a single subject or with clustered data (where individuals are grouped in clusters such as schools, clinics, etc.). In this context of data dependency, the Logistic regression is generalized by adding a random effect to account for the correlation between measurements of the same cluster (note that a subject can be also considered to be a cluster).

The mixed-effects Logistic regression is important because it sets the foundation for models with ordinal or model dependent variables, which can be seen as a generalization of the Logistic regression.

Let i indicate the level-2 units (subjects or clusters) and j the level-1 units (occasions or nested units). Also let $i = 1, \cdots, N$ be the level-2 units (e.g. subjects), and within each level-2 unit are nested $j = 1, \cdots, n_i$ level-1 units. Therefore, the total number of level-1 units across all clusters is computed as $n = \sum_{i=1}^{N} n_i$. Let Y_{ij} equal the value of the binary outcome, which can take 1 if the event of interest is present and 0 if not, corresponding to level-1 unit j nested within level-2 unit i.

The Logistic regression model can be expressed in terms of Logit and can include a random intercept as follows:

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \varsigma_i$$

where $\boldsymbol{x}_{ij} = (1, x_{ij1}, \cdots, x_{ijp})'$ denotes the $(p+1)$ by 1 vector of covariates for level-1 unit j within cluster, $\boldsymbol{\beta}$ the corresponding $(p+1)$ by 1 vector of regression coefficients, and ς_i is the random intercept specific to each level-2 unit (or cluster). These random effects are assumed to be $N(0, \sigma_\varsigma^2)$.

Expressing the random effects in standardized form, the model can be expressed as:

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma_\varsigma\theta_i$$

where θ_i is the standardized random intercept or $\theta_i = \frac{\varsigma_i}{\sigma_\varsigma}$.

The same model in terms of the latent variable y can be expressed as:

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sigma_\varsigma\theta_i + e_i$$

Looking at this formulation, it is easy to understand that the regression coefficients obtained from the mixed-effects model will differ from the ones obtained from the fixed-effects model. While the conditional variance of y given as set of covariates $\boldsymbol{x}$ equals $\sigma_\varsigma^2 + \sigma_e^2$, the conditional variance equals σ_e^2 in the fixed-effects model.

The estimates obtained from the mixed-effects model are usually called “subject-specific” since they are conditional on the random effect (note that subject is the level-2 unit). On the other hand, estimates from the fixed-effects or GEE models are termed population averaged estimates (i.e. marginal), indicating that the effect of a predictor is averaged over the population of individuals.

The *intraclass* correlation, which denotes the unexplained variance due to differences between subjects, equals $\frac{\sigma_\varsigma^2}{\sigma_\varsigma^2 + \frac{\pi^2}{3}}$ in the mixed-effects regression model.

Multilevel formulation – Let us assume a model with one level-1 variable x_{ij} and a level-2 variable x_i. Notice that in the case of repeated measure data, level-1 covariates are called time variant and level-2 covariates time invariant.

The level-1 model can be expressed in terms of the log odds as:

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0i} + \beta_{1i}x_{ij}$$

The level-2 model, assuming a random intercept only, is written as:

$$\beta_{0i} = \beta_0 + \beta_2 x_i + \varsigma_{0i}$$

$$\beta_{1i} = \beta_1$$

and assuming a random coefficient for x_{ij} (i.e. allowing a subject-specific slope for covariate x_{ij}) level-2 model is written as:

$$\beta_{0i} = \beta_0 + \beta_2 x_i + \varsigma_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_3 x_i + \varsigma_{1i}$$

The respective reduced forms, which are obtained by substituting the level-2 model in the level-1 model, are:

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \underbrace{\beta_0 + \ \beta_1 x_{ij} + \beta_2 x_i}_{fixed} + \underbrace{\varsigma_{0i}}_{random}$$

and,

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \underbrace{\beta_0 + \ \beta_1 x_{ij} + \beta_2 x_i + \beta_3 x_{ij} x_i}_{fixed} + \underbrace{\varsigma_{0i} + \varsigma_{1i} x_{ij}}_{random}$$

Following this formulation, these models can be easily generalized to include multiple level-1 or level-2 covariates.

Finally, one should take into account that, contrary to the multilevel models for continuous outcomes, the level-1 variance in the mixed-effects Logistic regression is fixed and not estimated.

One of the implications of this is that the level-1 variance cannot be reduced by adding level-1 covariates.

3.2.3.1.1.2. Method: Mixed-Effects Poisson Regression Model

Let us assume that y is the number of events occurring in a time interval of length t and is said to follow a Poisson distribution. The incidence rate is denoted by λ and the expectation of y is given by $\mu = \lambda t$. The Poisson regression is frequently utilized in modeling count data.

Fixed-effects Poisson regression – When counts are observed for different subjects i, the expectation of y_i can be modeled using a log-linear model as:

$$\mu_{ij} = E(y_i|\boldsymbol{x}_i') = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$$

or written as an additive log-linear model:

$$\ln(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta},$$

where y_i is a non-negative integer value or count variable (e.g. number of doctor visits) with expectation μ_{ij} given covariates $\boldsymbol{x}_i' = (1, x_{i1}, \cdots, x_{ip})'$. Notice that for the sake of simplicity this model assumes that the length of interval t is the same for all individuals. The exponentiated regression coefficients of a Poisson model are interpreted as rate ratios, or ratios of expected counts. A desirable property of this type of modeling is that if t is the same for all subjects, the exponentiated coefficient $\exp(\beta_k)$ can be interpreted as the incidence-rate ratio for a unit increase in covariate x_{ik}.

Random Intercept Poisson regression – In the ordinary Poisson regression, the independence assumption is not met when data are longitudinal or clustered. In such cases, the multilevel or mixed-effects models offer an alternative to address the dependence of the observations within subjects/clusters.

Let us assume that a total of $n = \sum_i n_i$ nonnegative integer values y_{ij} are measured for $i = 1, \cdots, N$ subjects and $j = 1, \cdots, n_i$ observations for subject i. The vector of covariates is denoted by $\boldsymbol{x}_{ij} = (1, x_{ij1}, \cdots, x_{ijp})'$ and $\boldsymbol{\beta}$ represents the corresponding vector of regression coefficients.

The random intercept Poisson model includes a subject-specific random intercept ς_i to account for the clustering in the data:

$$\begin{aligned}
\mu_{ij} = E(y_{ij}|\boldsymbol{x}'_{ij}, \varsigma_i) &= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \varsigma_i) \\
&= \exp(\beta_0 + \beta_1 x_{ij1} + \cdots + \beta_p x_{ijp} + \varsigma_i) \\
&= \exp(\varsigma_i)\exp(\beta_0 + \beta_1 x_{ij1} + \cdots + \beta_p x_{ijp})
\end{aligned}$$

where ς_i is assumed to follow $N(0, \sigma_\varsigma^2)$.

The model where the random effect is expressed in standardized form is written as:

$$\mu_{ij} = E(y_{ij}|\boldsymbol{x}'_{ij}, \varsigma_{0i}) = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\varsigma \theta_i)$$

where θ_i is the standardized random intercept or $\theta_i = \frac{\varsigma_i}{\sigma_\varsigma}$.

3.2.3.2. Review: Heterogeneity Models

The mixed models covered up to this point assume normality of the random effects, which implies that they are sampled from a homogeneous population of random effects. However, it has been shown that misspecification of random effects distribution can lead to biased parameter estimates. Therefore, methods that relax this distributional assumption for the random effects are necessary (Molenberghs & Verbeke, 2005). Substituting the normality assumption of the random effects by a mixture of normal distributions leads to the heterogeneity approaches (Verbeke & Molenberghs, 2009; Molenberghs & Verbeke, 2005). This method assumes that the random effects are drawn from a mixture of normal distributions, not just one single normal distribution, which reflects the assumed presence of unobserved heterogeneity (Verbeke & Molenberghs, 2009; Molenberghs & Verbeke, 2005). Several advantages arise from the heterogeneity model:

1. Using the finite mixtures of normal distributions will result in a flexible model;
2. The mixture distributions can be utilized to model unobserved heterogeneity in the random effects;
3. The mixture distributions can be used for the purpose of classification and hence are useful for cluster/discriminant analysis for longitudinal data (Verbeke & Molenberghs, 2009; Molenberghs and Verbeke, 2005).

Despite the usefulness of this model, this review found no applications of heterogeneity models in Education and Psychology research during the past 9 to 10 years. It remains a useful tool for future research.

3.2.3.2.1. Method: The Heterogeneity Model

The heterogeneity model was introduced by Verbeke and Lesaffre (1996) as an extension of the LMMs to cases in which the distribution of the random effects was not from a single normal distribution. This model is an accepted method for classifying longitudinal profiles.

Let us consider the following LMM:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varsigma_{0i} + \varsigma_{1i} t_{ij} + e_{ij}$$

Making a small change in the notation of the random-effects, one can rewrite it as:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

This model can be generalized to a LMM with more covariates and random-effects, and expressed in matrix form as:

$$\boldsymbol{Y_i} = X_i \boldsymbol{\beta} + Z_i \boldsymbol{b_i} + \boldsymbol{e_i}$$

where,

- $\boldsymbol{b_i}$ represents the vector of random effects,
- $\boldsymbol{\beta}$ represents the vector of parameters corresponding to the fixed-effects part of the model.

Substituting the normality assumption of the random-effects by a mixture of K, q-dimensional normal distributions will result in a heterogeneity model with random effects that have mean vectors $\boldsymbol{\mu_k}$ and covariance matrices D_k as follows:

$$\boldsymbol{b_i} \sim \sum_{k=1}^{K} p_k N(\boldsymbol{\mu_k}, D_k)$$

where $\sum_{k=1}^{K} p_k = 1$. The additional constraint $\sum_{k=1}^{K} p_k \, \boldsymbol{\mu_k} = 0$ is needed to ensure that $E(\boldsymbol{y_i}) = X_i\beta$. One can assume that all covariance matrices are the same, $D_k = D$ for all k. Notice that k denotes group membership to latent group or class k.

The heterogeneity model is then specified as:

$$\boldsymbol{Y_i} = X_i\beta + Z_i\boldsymbol{b_i} + \boldsymbol{e_i},$$

$$\boldsymbol{b_i} \sim \sum_{k=1}^{K} p_k N(\boldsymbol{\mu_k}, D),$$

$$\textstyle\sum_{k=1}^{K} p_k = 1, \quad \sum_{k=1}^{K} p_k \, \boldsymbol{\mu_k} = 0$$

$$\boldsymbol{e_i} \sim N(\boldsymbol{0}, \Sigma_i)$$

$$b_1, \cdots, b_N \; and \; e_1, \cdots, e_N \; are \; independent$$

This model assumes that the population of random-effects consists of a mixture of K subpopulations or latent classes with mean vectors $\boldsymbol{\mu_k}$ and covariance matrix D. Note that in comparison, the LMM assumed that the random effects had mean zero, $\boldsymbol{b_i} \sim N(0, D)$.

To illustrate the heterogeneity model with a simple example, let us go back to the LMM with equation:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

and let us assume that the random coefficient for time, b_{1i}, has different means across two latent classes or subpopulations. Therefore, b_{1i} will no longer follow a $N(0, \sigma_{\varsigma}^2)$, but it will follow a mixture of two normal distributions as follows:

$$b_{1i} \sim p_1 N\left(\mu_1, \sigma_{1\varsigma}^2\right) + p_1 N\left(\mu_2, \sigma_{2\varsigma}^2\right),$$

where μ_1, μ_2 and $\sigma_{1\varsigma}^2$, $\sigma_{2\varsigma}^2$ (note that one can have $\sigma_{1\varsigma}^2 = \sigma_{2\varsigma}^2$) represents the means and variances of the random coefficient b_{1i} for each subpopulation, respectively, and where p_1 is the unknown proportion of subjects in subpopulation 1 in the dataset and $p_2 = 1 - p_1$ is the proportion of subjects in subpopulation 2.

Notice that since the covariate in this model is time, this mixture model also receives the name of *mixture growth model*. One of the goals of the mixture growth model is identifying clusters of individuals with similar growth parameters.

Model estimation: the EM algorithm – The marginal distribution of $\boldsymbol{Y_i}$ under the heterogeneity model is given by:

$$\boldsymbol{Y_i} \sim \sum_{k=1}^{K} p_k N(X_i \beta + Z_i \boldsymbol{\mu_k}, V_i)$$

where $V_i = Z_i D Z_i' + \Sigma_i$. The estimation of the parameters β, $\boldsymbol{\mu_k}$, p_k, D and the parameters in Σ_i can be done using maximum likelihood estimation. In this context of mixture problems, the Expectation-Maximization (EM) algorithm is very useful since it can happen that when a model is fitted with too many parameters due to having a g too large, the likelihood function can be maximal anywhere on a ridge and, therefore, not able to find a solution. The EM algorithm can find convergence in some particular point on that ridge.

Let us assume the following specifications:

- $\boldsymbol{\pi}$ represents the vector of probabilities, i.e., $\boldsymbol{\pi}' = (p_1, \cdots, p_K)$,
- $\boldsymbol{\gamma}$ represents the vector comprising the parameters β, D, the covariance components in Σ_i and the means in $\boldsymbol{\mu_k}$,
- $\boldsymbol{\theta}' = (\boldsymbol{\pi}', \boldsymbol{\gamma}')$ represents the vector of all parameters in the marginal heterogeneity model,
- $f_{ik}(\boldsymbol{y_i} | \boldsymbol{\gamma})$ represents the density function of $N(X_i\beta + Z_i\mu_k, V_i)$.

The corresponding likelihood function can be expressed as:

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{N} \left\{ \sum_{k=1}^{K} p_k f_{ik}(\boldsymbol{y_i} | \boldsymbol{\gamma}) \right\}$$

where $\boldsymbol{y}' = (\boldsymbol{y_1'}, \cdots, \boldsymbol{y_N'})$ is the vector that contains the measured outcome variables for the N subjects.

Let us define z_{ik} such that the prior probability that a subject belongs to component or group k is $P(z_{ik} = 1) = p_k$. The log-likelihood for the observed $\boldsymbol{y}$ and for the vector $\boldsymbol{z}$ of the unobserved indicators z_{ik} can be written as:

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \{\ln p_k + \ln f_{ik}(\boldsymbol{y_i} | \boldsymbol{\gamma})\}$$

While maximizing $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ is easier than maximizing $\ell(\boldsymbol{\theta}|\boldsymbol{y})$, it yields estimates which depend on the unobserved z_{ik}. The EM algorithm solves this problem by maximizing the conditional expectation of $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ instead of $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ itself. In the expectation step (i.e. E), the conditional expectation of $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ given the observed vector $\boldsymbol{y}$ is computed. Then, in the maximization step (i.e. M), the conditional expectation of $\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$ is maximized with respect to $\boldsymbol{\theta}$, and finally an updated estimate for $\boldsymbol{\theta}$ is recorded. The above algorithm iterates between the steps E and M until convergence is achieved.

3.2.4. Review: Conditional Models/Transition Models

The parameters of conditional approaches describe a feature of a set of responses given certain values for the other responses (Cox, 1972). The mentioned feature could be odds, logits, probabilities, and so on. Thus, conditional models refer to methods that model the mean and time dependence simultaneously by means of conditioning a response variable on other responses (or on a subset of other responses) (Diggle et al., 2002; Molenberghs & Verbeke, 2005; Fitzmaurice & Molenberghs, 2009). Diggle at al. (2002, pp. 142–144) have criticized the use of conditional models since these methods have difficulty interpreting the fixed-effect parameters, such as the treatment effect of one outcome as it is modeled conditionally on other outcomes for the same participant, the responses of other participants, and the number of repeated measures. (Note that when adding or deleting an observation for an individual, the value and interpretation of the parameter would change.)

Nonetheless, a particular case of the conditional models is so-called transition, or Markov, models (P. Diggle, Diggle, et al., 2002; Fitzmaurice & Molenberghs, 2009; Molenberghs & Verbeke, 2005). Transition models are useful for repeated measure data because in such approaches the conditional distribution of each outcome is written as an explicit function of the previous outcomes and variables (P. Diggle, Diggle, et al., 2002; Fitzmaurice & Molenberghs, 2009; Molenberghs & Verbeke, 2005). Transition approaches can be thought of as conditional models where one models the conditional distribution of the dependent variable at any time point given the past outcomes and variables. Transition or Markov models are a specific type of conditional model which accounts for dependence among the repeated measures by conditioning a response variable on the other responses of the same subject that allows the past measurements to influence the present values of the subject (P. Diggle, Diggle, et al., 2002). The most useful transition models are Markov chains for discrete response variables, where the conditional distribution of each response depends on the q prior observations (P. Diggle, Diggle, et al., 2002; Fitzmaurice & Molenberghs, 2009). The integer q represents the order of a transition model, i.e., the number of past measurements that can influence the current one.

Transition models have their own limitations for the analysis of repeated measure data. For instance, transition models require time points that are equally spaced in time, and hence it is difficult to utilize transition models when data are missing or when intervals between repeated measures are irregularly spaced (Fitzmaurice & Molenberghs, 2009). Note that the interpretation of parameters changes with the order of the serial dependence (P. Diggle, Diggle, et al., 2002; Fitzmaurice & Molenberghs, 2009). Finally, conditioning on the past outcomes may lessen the effects of variables of interest (Fitzmaurice & Molenberghs, 2009). Regardless, transition (Markov) models have been utilized for analyzing repeated measure data in Education and Psychology; see, for example, Berridge, Penn and Ganjali (2009); Facal, Guàrdia-Olmos, and Juncos-Rabadán (2015); and Allik and Kearns (2017).

3.2.4.1. Method: Conditional Linear Mixed Models

In repeated measure data, particularly in observational studies, baseline differences between subjects need to be taken into account. In other words, the longitudinal changes need to be corrected for potential confounders such as age, gender, etc. These subject characteristics are called the cross-sectional component, which is usually treated as a nuisance since the primary interest when analyzing repeated measure data is generally the longitudinal or time effects.

Let us consider the following LMM:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 t_{ij} + \beta_4 x_{1i} t_{ij} + \varsigma_{0i} + \varsigma_{1i} t_{ij} + e_{ij}$$

where,

- β_0 is the grand mean,
- x_{1i} is a time-invariant covariate (e.g. age at study entry),
- t_{ij} denotes the time point at which the jth measurement from subject i is taken,
- ς_{0i} represents the random intercept,
- ς_{1i} the random coefficient or slope for time.
- e_{ij} represents the measurement error.

One can rearrange the equation in terms of the cross-sectional and longitudinal components as follows:

$$y_{ij} = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \varsigma_{0i}) + \ (\beta_3 + \beta_4 x_{1i} + \varsigma_{1i}) t_{ij} + e_{ij}$$

Then, to simplify notation we will replace ς_{0i} and ς_{1i} by b_{0i} and b_{1i}, respectively:

$$y_{ij} = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + b_{0i}) + \ (\beta_3 + \beta_4 x_{1i} + b_{1i}) t_{ij} + e_{ij}.$$

Misspecifications in the cross-sectional component, such as omitting a relevant covariate, can carry negative effects for the estimates of the longitudinal effects, including inflated random-intercept variance and bias in the estimation of the random effect covariance.

Verbeke and Lesaffre (1999) and Verbeke, Spiessens and Lesaffre (2001) introduced the *conditional LMMs* as an alternative to analyzing repeated measure data without the need to specify any cross-sectional variables.

The LMM above can be reformulated as:

$$y_{ij} = b_i^* + \ (\beta_3 + \beta_4 x_{1i} + b_{1i}) t_{ij} + e_{ij}$$

where,

- The parameters of interest are fixed slopes β_3, β_4,
- b_i^* represents the cross-sectional component for subject i under the original model,
- The subject-specific slope b_{1i},

- The residual variance σ^2.

Notice that in this model, the cross-sectional effect b_i^* is considered a nuisance. This model can be expressed using the following form:

$$\boldsymbol{Y}_i = \mathbf{1}_{n_i} b_i^* + X_i \beta + Z_i \boldsymbol{b_i} + e_i$$

where the matrices X_i, Z_i and vectors β and b_i correspond to the sub-matrices and sub-vectors of their original counterparts obtained by deleting those elements corresponding to the cross-sectional effect (i.e. the time-invariant variables) from the original presentation of the method.

Model estimation – Estimating the conditional LMMs consists of two steps. The first step of the estimation consists of conditioning on sufficient statistics for the nuisance parameters b_i^*. The second step is utilizing the maximum likelihood (ML) or the restricted maximum likelihood estimation (REML) to estimate the rest of the parameters using the conditional distribution of the vector of responses Y_i on the sufficient statistics.

Let us consider $\bar{y}_i = \sum_j \frac{y_{ij}}{n_i}$ to be a sufficient statistic for b_i^*. The distribution of $\boldsymbol{Y_i}$, conditional on $\bar{y}_i$ and the subject-specific random effects $\boldsymbol{b_i}$, can be written as:

$$f_i(\boldsymbol{y_i}|\bar{y}_i, \boldsymbol{b_i}) = \frac{f_i(\boldsymbol{y_i}|b_i^*, \boldsymbol{b_i})}{f_i(\bar{y}_i|b_i^*, \boldsymbol{b_i})}.$$

After some matrix algebra, it follows that the distribution above is proportional to:

$$(2\pi\sigma^2)^{-\frac{n_i-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(A_i'\boldsymbol{y_i} - A_i'X_i\beta - A_i'Z_i\boldsymbol{b_i})'(A_i'A_i)^{-1}\right.$$
$$\left.\times (A_i'\boldsymbol{y_i} - A_i'X_i\beta - A_i'Z_i\boldsymbol{b_i})\right\}$$

for any set of $n_i \times (n_i - 1)$ matrices A_i of rank $n_i - 1$ which satisfy $A_i'\mathbf{1}_{n_i} = 0$. If one also adds the condition that $A_i'A_i = I_{n_i-1}$, then the conditional approach is equivalent to estimating the transformed model:

$$\boldsymbol{Y_i^*} \equiv A_i'\boldsymbol{Y}_i = A_i'X_i\beta + A_i'Z_i\boldsymbol{b_i} + A_i'e_i$$

$$= X_{\boldsymbol{i}}^{*}\beta + Z_{\boldsymbol{i}}^{*}\boldsymbol{b_i} + e_i^{*},$$

where $X_{\boldsymbol{i}}^{*}$, $Z_{\boldsymbol{i}}^{*}$ and e_i^{*} are $N(0, \sigma^2 I_{n_i-1})$. One can see that the only parameters in the transformed model shown above are the longitudinal effects and the residual variance. This transformed LMM can be estimated via ML or REML methods.

The simplest case of a conditional LMM is that of balanced data with only two measurements per participant; here the only time-variant variable of interest is the occasion in which the measurement was taken (pre, post). The conditional model in this context is the same as analyzing the difference between pre and post for each participant. Therefore, conditional LMMs can be seen as an extension of the paired t-test to more than two measurements per subject and unbalanced data.

The main advantage of conditional LMMs is that the inference is simpler than with traditional LMMs because the first conditional step decreases the complexity of the algorithms for model fittings. Also, the second step of the approach lifts the normality assumption for the random longitudinal effects. A disadvantage of this method is that all subject-specific cross-sectional information from the variables is lost.

3.2.4.2. Method: The Transition Model

Transition models belong to the family of conditional models. Conditional models can be understood as the extension of GLMs to the case of repeated measure data. This is done by modeling the time dependency simultaneously by conditioning a response on other responses. In other words, outcomes are modeled conditionally on the value of a subset of other responses of the same cluster. In the longitudinal data context such a subset can either include all measurements taken previous to the measurement being modeled, or only the most recent measurements. These models are called *transition models*.

For transition models, the conditional distribution of the outcome at any time point is modeled using the past outcomes and variables. Therefore, the dependence/correlation among the repeated measures of a subject is assumed to be due to the influence of past measurements of the response in the present value. This model can be written as:

$$g^{-1}\{E(Y_{ij}|\boldsymbol{X}_{ij},\boldsymbol{H}_{ij})\} = \boldsymbol{X}'_{ij}\beta + \sum_{r=1}^{s}\alpha_r f_r(\boldsymbol{H}_{ij}),$$

where $\boldsymbol{H}_{ij} = (Y_{i1},\cdots,Y_{ij-1})$ represents the vector of past responses corresponding to the jth occasion; $f_r(\boldsymbol{H}_{ij})$ represents a function, not necessarily linear, of the past responses; and g denotes the link function (already introduced in the GEE models).

Note that a particular case of transition models is the first-order autoregressive (i.e. $AR(1)$), GLM, which is obtained when:

$$\sum_{r=1}^{s}\alpha_r f_r(\boldsymbol{H}_{ij}) = \alpha_1 f_1(\boldsymbol{H}_{ij}) = \alpha_1 Y_{ij-1}$$

In this model, the current outcome is assumed to depend only on the previous response besides the variables and the link function is the identity. The $AR(1)$ model can be expanded to an autoregressive of order s, $AR(s)$, where the s previous responses are taken into account,

$$\sum_{r=1}^{s}\alpha_r f_r(\boldsymbol{H}_{ij}) = \alpha_1 Y_{ij-1} + \cdots + \alpha_s Y_{ij-s}$$

In general, these types of conditional models, where the outcome in occasion j depends only on the s prior outcomes, are called Markov models of order s. More specifically, these models are called *Markov chain models* for discrete outcomes.

Markov chain models have been extensively used in discrete repeated measures that are equally spaced with a finite number of states $S = \{s_1, s_2, \cdots, s_r\}$. The process starts in one of these states and moves successively from one state to another. The transition probabilities, p_{ij} represent the probability that the chain currently in state s_i moves to state s_j. Note that, the conditional probability of going into each state, given the previous state, is called the transition probability.

The process can also remain in the current state and the probability associated with this event is p_{ii}. Usually a particular state is specified as the starting state.

In the simplest case which is a first-order Markov chain, the model can be expressed in terms of the initial state and a set of transition probabilities; they are assumed to be the same for each time interval. Note that since the model is first order, dependence is present only on the immediately previous state.

In more general models, one can incorporate higher orders to implement dependence on more than the immediately previous measurements. Additionally, the transition probabilities are permitted to change across occasions.

An attractive feature of transition modeling is the following presentation for the conditional likelihood, given a set of s initial measurements; note that the joint distribution of the outcome vectors is written as a product of conditional distributions:

$$f(Y_{i1}, \cdots, Y_{in}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{j=1}^{n} f(Y_{ij} | Y_{ij-s}, \cdots, Y_{ij-1}; \boldsymbol{\beta}, \boldsymbol{\alpha}),$$

Though Markov chain and autoregressive models have been applied to longitudinal data, there are some limitations. First, transition models assume that the time points are equally distant in time. Second, the presence of missing measurements complicates the use of transition models. Finally, these models are not recommended when the target of inference is the regression parameters $\boldsymbol{\beta}$ since their estimation is very sensitive to the order of dependence established in the model. In other words, incorporating the history of past responses may attenuate the effect of covariates of interest.

3.2.5. Review: Structural Equation Modeling (SEM) Approaches

SEM approaches are utilized for *specifying* and *estimating* models of the nature and strength of the relationships among observed and latent variables (i.e., theoretical constructs that cannot be directly measured) (MacCallum & Austin, 2000). There are two main components in SEM: a *measurement* model and *structural* model. A measurement model describes and estimates the associations between the observed variables and the latent constructs. A structural model defines and estimates the hypothesized relationships between the latent constructs themselves (Little, 2013). SEM is heavily used to analyze longitudinal data in studies of developmental psychology and aging (MacCallum & Austin, 2000). There are two core approaches in modeling repeated measure data within the SEM framework, namely, autoregressive and latent growth curve approaches (Bollen & Curran, 2004, 2006). Both methods have the following advantages of SEM (Bollen & Curran, 2004, 2006):

1. They estimate and correct for time-specific measurement error,
2. They use multiple indicators per latent variable,
3. They model mediating and moderating effects,
4. They explore measurement invariance across time.

However, because SEM is considered to be an independent set of statistical techniques (when compared to ANOVA or regression-like models), the statistical methods related to SEM approaches will not be reviewed here. Hoyle (1995) and Kline (2015) are two great resources for readers interested in the statistical details of SEM approaches.

3.2.5.1. Review: SEM Autoregressive Models

Autoregressive models utilize the same variable's previous observation to predict its future measurement. In other words, for an autoregressive approach that has one latent variable, the value of the latent variable at each occasion is mainly influenced by (and therefore modeled on) the value of that same latent variable at the previous occasion (Selig & Little, 2012). An autoregressive cross-lagged model could be utilized for studies with more than one latent variable. In this case, it would model the unique effect of one variable measured at an earlier time point (e.g. Time 1) on another variable observed at a later time point (e.g. Time 2), while accounting for the autoregressive effects of that variable at Time 1 on itself at Time 2 (Selig & Little, 2012). Autoregressive models test how between-person differences in levels of a variable at one occasion are predicted by between-person differences in the same variable at a previous occasion based on a SEM framework (Selig & Little, 2012). The autoregressive models are particularly useful when the purpose of the study is to detect the relations between variables over time and to examine the direction of causation over time (Selig & Little, 2012). The models are also well suited to repeated measure data strings with sequential transmission; when values of the variable of interest at Time 3 rest on the values at Time 2, which in turn rest on the values at Time 1 (Selig & Little, 2012). Brock, Nishida, Chiong, Grimm and Rimm-Kaufman (2008) had conducted a series of auto-regressive cross-lagged models (using SEM context) to examine the relation between responsive classroom teacher practices, children's perceptions, and their academic and social competence (in terms of social skills, academic performance, standardized reading and math scores), using data collected over a 3-year period from 520 children attending Grades 3-5 in one of six chosen schools in the northeast United States. More recent applications of autoregressive (cross-lagged) models in the Educational and Psychological research can be seen in Abbott, Berninger and Fayol (2010); Hong, Yoo, You and Wu (2010); Li and Lerner (2013); Guo, Sun, Breit-Smith, Morrison and Connor (2015); and Ciarrochi et al. (2016).

3.2.5.2. Review: SEM Latent Growth Curve Models

Latent growth curve models using the SEM framework allow scholars to model and estimate:

- A mean growth trajectory for a cluster of interest on an outcome,
- Variability in subject patterns around the mean pattern,
- The degree to which certain time-invariant and time-varying variables predict subject variability (Curran, Obeidat, & Losardo, 2010).

The process of latent growth curve modeling consists of two stages. In the first stage, parameters of individual growth patterns, such as intercepts and slopes along with differences in these patterns, are described and estimated. In the second stage, predictors are utilized to describe the variance in individual growth patterns (Bollen & Curran, 2006). In other words, latent growth curve modeling permits the estimation of inter-individual variance in intra-individual structure of change across time (Bollen & Curran, 2006). Questions that can be addressed by latent growth curve modeling include:

1. Concerns regarding the characteristics such as rate and shape of the overall pattern of the sample,
2. The influence of different predictors on the variability of individual growth trajectories,
3. Individual differences in trajectories (Bollen & Curran, 2006).

The primary advantages of latent growth curve models using the SEM framework are the models' high flexibility; they can incorporate a variety of complexities such as partly missing data, unequally spaced time measure, time-varying covariates, complex non-linear or compound-shaped trajectories, non-normally distributed or discrete repeated measures, and multivariate growth processes (Curran et al., 2010). Furthermore, many simulation studies have shown that latent growth curve models usually have higher statistical power compared to similar traditional methods (Curran et al., 2010). Applications of the latent growth curve models under the SEM framework are widely adopted in the literature. For example, Brailean et al. (2017) used a series of latent growth curve models, as a function of time, for each dependent variable including but not limited to depressed affect, positive affect, immediate recall, and inductive reasoning. This allowed the authors to examine:

1. The intercept, i.e. the initial level of a specific response,
2. The slope, i.e. the rate and form of change (which could be linear or non-linear latent growth trajectories),
3. The relation between the intercept and slope.

Other examples of the applications of latent growth curve models under the SEM framework in Education and Psychology include Caprara et al. (2008); Simons-Morton and Chen (2009); Mäkikangas, Bakker, Aunola and Demerouti (2010); Ng, Feldman and Lam (2010); King (2015); Ciarrochi et al. (2016); and Ladd, Ettekal, and Kochenderfer-Ladd (2017). As mentioned before, SEM approaches are distinct from RM ANOVA and LMM, so the methodological details of SEM Latent Growth Curve modeling will not be covered here.

3.2.6. Review: Mixture Models for Longitudinal Data

For mixed-effects models with a single-component multivariate normal distribution (for more details see Section 3.2.2.), the assumption of random effects is that the participants originate from a homogeneous population and can therefore be described by one mean and variance-covariance structure (Fitzmaurice et al., 2004). However, this assumption might not be realistic when different subpopulations of participants exist, each with its own pattern (Molenberghs & Verbeke, 2005). Mixture models have been widely used in many disciplines due to their ability to capture the subgroup heterogeneity (Tang & Qu, 2016). Finite-mixture approaches can be viewed as latent-variable methods that can model the distribution of a variable as a mixture of a finite number of distributions (McLachlan & Peel, 2000).

Mixture models have been utilized for repeated measure data in different setups. For example, in the heterogeneity models discussed in Section 3.2.2.2, heterogeneity is allowed by relaxing the random-effects normality assumption to incorporate mixtures of normal components (Molenberghs & Verbeke, 2005; Verbeke & Molenberghs, 2000). Other examples of mixture modeling for repeated measure data are the latent class growth model (LCGM) and the latent growth mixture model (LGMM) (Jung & Wickrama, 2008; Vermunt, 2010; Berlin, Williams, & Parra, 2014). The primary goals of LCGM and LGMM are:

1. To understand subject variability in parameters reflecting individual change in the dependent variable across measurement occasions,
2. To probabilistically allocate subjects into subpopulations; this can be done by assigning each subject to a latent class, where the observed distribution of measurements may be a mixture of two or more subpopulations (Berlin et al., 2014).

Although the LCGM and LGMM are closely related, the main distinction between them is the values which are permitted to differ within and between latent classes (Berlin et al., 2014; Jung & Wickrama, 2008). LGMM allows researchers to control which parameters can vary both within and between classes; these parameters include latent variables' means, variances, covariances, residuals, and so on. On the other hand, in LCGM, the variance of latent slope and intercept within each class are fixed to zero, but they are permitted to vary between classes (Berlin et al., 2014; Jung

& Wickrama, 2008). Therefore, fewer parameters need to be estimated for LCGM than for LGMM. Additionally, LCGM assumes that all subject growth patterns are homogeneous within classes (Berlin et al., 2014; Jung & Wickrama, 2008). Besides these three applications of mixture models (heterogeneity models, LCGM, and LGMM), researchers have proposed different methods to incorporate mixture models into longitudinal data analysis. For example, Muthén and Shedden (1999) proposed a variation of mixture modeling that permits the joint estimation of the following:

1. A Logistic regression of binary outcomes on classes,
2. A finite-mixture growth model where distinct shaped curves are presented by class varying random-coefficient means.

Sun, Rosen, and Sampson (2007) proposed a multivariate Bernoulli mixture model by employing random effects for mixing proportion in the GLM framework.

Mixture models have been utilized in Education and Psychology research for repeated measure data. For example, Hart, Musci, Slemrod, Flitsch and Ialongo (2018) fitted a latent class growth model to explore the developmental patterns of aggressive-disruptive symptoms. Ladd et al. (2017) performed latent growth mixture modeling to classify children with similar victimization patterns from kindergarten to 12th grade.

There exist three types of mixture models for repeated measure data, namely, the mixture growth, Mixture Markov, and latent Markov models. Only the Mixture Markov model will be covered below.

3.2.6.1. Method: The Mixture Markov Model

In section 3.2.4., the transitional model was introduced, and the case of the first-order Markov model was discussed. This model assumes that Y_{ij} depends *only* on Y_{ij-1}. One of the limitations of such a model is that it assumes that the transition probabilities are homogeneous. The mixture Markov model is introduced to take into account unobserved heterogeneity, i.e. to allow transition probabilities to differ across unobserved subgroups of subjects.

The first-order mixture Markov model can be expressed as:

$$f(Y_{i1},\cdots,Y_{in}) = f(\boldsymbol{Y_i}) = \sum_{\boldsymbol{\ell=1}}^{L} P(\,w_i = \boldsymbol{\ell})f(Y_{i0}|w_i = \boldsymbol{\ell})\prod_{j=1}^{n} f\big(Y_{ij}\big|\,Y_{ij-1},w_i = \boldsymbol{\ell}\big)$$

where L denotes the $\boldsymbol{\ell} = 1,\cdots,L$ latent classes. These latent classes are assumed to vary regarding the initial-state and transition densities. The probability $P(w_i = \boldsymbol{\ell})$ indicates that subject i belongs to class $\boldsymbol{\ell}$**.**

For categorical response variables, the model can be expressed as:

$$P(\boldsymbol{Y_i}) = \sum_{\boldsymbol{\ell=1}}^{L} P(\,w_i = \boldsymbol{\ell})P(Y_{i0} = m_0|w_i = \boldsymbol{\ell})\prod_{j=1}^{n} P\big(Y_{ij} = m_j\big|\,Y_{ij-1} = m_{j-1},w_i = \boldsymbol{\ell}\big)$$

Special cases of the mixture Markov model – By implementing restrictive conditions on the transition probabilities one can obtain various special cases of the mixture Markov approach. For example, a “mover-stayer” model (Goodman, 1961) is a two-class model ($L = 2$) characterized by the fact that subjects in one of the classes - for instance, the second - have zero probability of making a changeover. In other words:

$$P\big(Y_{ij} = m_j\big|Y_{ij-1} = m_{j-1},w_i = 2\big) = 0 \quad \text{for } m_j \neq m_{j-1}.$$

Another example entails a Markov model in which the measurements for a random latent class are independent across occasions, that is:

$$P\big(Y_{ij} = m_j\big|Y_{ij-1} = m_{j-1},w_i = 2\big) = P\big(Y_{ij} = m_j\big|w_i = 2\big).$$

In other words, the probability of a subject’s response on a particular occasion is independent of the previous response if the subject belongs to, say, the second latent class.

Extensions of the mixture Markov model – The most common extension of the simple mixture Markov model incorporates independent variables that affect class membership, the initial state, and transition probabilities. For instance, one can write regression models for Y_{i0} and Y_{ij} to include variables that affect the initial state and the transitions, respectively.

3.2.7. Review: Time-Series Approaches

Time-series analysis is an important analytical technique to understand and predict the behavior of variables in different disciplines (Jebb, Tay, Wang, & Huang, 2015). Unlike longitudinal data that frequently contain numerous measurements from many subjects, data from a time series contain several observations originating from very few sources, or just one. Furthermore, the length of time-series is generally at least 20 observations long, which is often longer than the length of longitudinal data, and to obtain precise estimation, many time series modeling approaches require 50 or more observations (McDowall, McCleary, Meidinger, & Hay, 1980, p. 20). Time series studies exhibit a unique structure that often demonstrates characteristics that are seldom observed in the repeated measure data typically collected in psychological research. Generally, time series studies have four components:

- Trend: Trend in time series data means any systematic long-term change/direction in the series level (Hyndman & Athanasopoulos, 2018; McDowall et al., 1980).
- Seasonality: The seasonal element of a time series is an increase/decrease trajectory that consistently reoccurs during the series. In other words, it is a cyclical or repeating structure of the measure within a certain time interval that is ascribed to seasonal aspects (Hyndman & Athanasopoulos, 2018).
- Cycles: A cycle in a time series is an isolating pattern such as increase or decrease that reoccurs over a certain time interval. Note that seasonal effects have fixed

intervals between occurrences which are related to some calendar feature, the patterns of cyclical effects do not have fixed intervals, which means that their length frequently differs from one cycle to another and cannot be attributed to any naturally occurring time periods (Hyndman & Athanasopoulos, 2018).

- Irregular variation (randomness): While trend, seasonality, and cycles all characterize systematic structure of variability in time series, there exist irregular components that characterize statistical noise and represent any leftover variation in a time series that was not accounted for with the three mentioned systematic components.

While performing any time-series analysis, if a systematic pattern (e.g. trend, seasonality, or cycles) has been observed, it must either be explicitly modeled or removed using transformations such as detrending or seasonal adjustments (Hyndman & Athanasopoulos, 2018). An effective statistical model accounts for all the mentioned systematic components (i.e. trend, seasonality, and cycles), translating the residuals to white noise (i.e. mean zero and constant variance).

In Education and Psychology research, the current observation may partly depend on its previous states, which means many educational/psychological variables often display autocorrelation (Jebb et al., 2015). Time-series approaches are designed to account for the effect of previous measurements by including this source of variance that might be potentially significant (Hyndman & Athanasopoulos, 2018). Essentially, time-series analysis assumes that the observations contain a systematic pattern along with a random noise that makes identifying this systematic pattern difficult (Hyndman & Athanasopoulos, 2018; Menard, 2002). Examples of the application of time-series analysis in the field of Education and Psychology include Webb, Sheeran and Luszczynska (2009); Smith, Handler and Nash (2010); Shelton, Hung and Baughman (2016); Markowitz (2018);

and Pennings et al. (2018). Due to the high technicality of time-series approaches, further discussion of this topic is not presented here. The references in this section provide useful resources for readers interested in the methodological details of time-series modeling.

3.2.8. Review: Covariance Structure Modeling

A primary limitation of mixed effects modeling is its dependence on correctly specifying the mean and correlation patterns of the repeated measure outcomes to ensure valid hypothesis testing and the correct conclusions (Fitzmaurice et al., 2004). Thus, it is crucial to consider methods for appropriately accounting for the correlation between the repeated measures from the same subjects (Fitzmaurice et al., 2004). Only after implementing an appropriate covariance structure can valid standard errors be estimated, which will result in correct inferences. Note that accounting for the correlation among repeated measures could increase efficiency of parameter estimation. In other words, not accounting for the correlation among the measurement occasions would cause incorrect estimates of the sampling variability; this could cause misleading inferences and consequently wrong conclusions (Wolfinger, 1993, 1996; Keselman, Algina, Kowalchuk, & Wolfinger, 1998; Littell, Pendergast, & Natarajan, 2000; Fitzmaurice et al., 2004; Kwok et al., 2007; Dedrick et al., 2009; Barnett, Koper, Dobson, Schmiegelow, & Manseau, 2010; Pusponegoro, Notodiputro, & Sartono, 2017).

An extensive number of covariance patterns has been introduced in literature, including but not limited to:

- Diagonal (I),
- Compound symmetry (CS),
- Variance components (VC),
- Banded (UN(2)),

- Toeplitz (TOEP),
- Banded Toeplitz (TOEP(2)),
- Compound symmetry with heterogeneous groups (CS*GROUP),
- First-order autoregressive (AR(1)),
- $AR(1)$ plus, diagonal ($AR(1) +$ I),
- $AR(1)$ plus, common covariance ($AR(1)$ + J),
- Spatial power law (SP(POW)),
- Unstructured (UN) (R. D. Wolfinger, 1996).

Wolfinger (1996) further provided a set of heterogeneous covariance patterns for longitudinal data that includes:

- Heterogeneous counterparts of the CS and $AR(1)$ structures,
- The independent-increments pattern,
- The first-order antedependence model,
- The Huynh-Feldt pattern,
- Correlated random coefficients models,
- A simplified factor-analytic construction.

Among the available covariance pattern models for accounting/explaining variability, the most common ones are I, CS, UN, and $AR(1)$ (Barnett et al., 2010; R. Wolfinger, 1993; R. D. Wolfinger, 1996). These correlation patterns can be defined as:

- The I covariance pattern assumes no correlation between observations, and it is used when none of the outcomes are correlated.
- The CS covariance pattern assumes that the correlation between any two observations is the same, regardless of the time lag between them.

- The $AR(1)$ covariance pattern assumes a stable decrease in correlation with increasing time lag or distance between measurements.
- The UN covariance pattern assumes that no two observations have the same correlation, and no structure is defined between adjacent measurements.

With the rising acknowledgment of the importance of the covariance pattern selection, methods have been established that permit researchers to make decisions about which covariance structure to apply according to the method used. Often the underlying correlation pattern of the repeated measure is not known in advance. Thus, researchers are forced to investigate various patterns and depend on fit criteria to choose among different conceivable covariance patterns (R. Wolfinger, 1993). Two common fit indices are Akaike's Information Criterion (AIC) (Akaike, 1974) and Schwarz's Bayesian Information Criterion (BIC) (Schwarz, 1978). Both the AIC and the BIC start with the loglikelihood function and penalize for the number of parameters to be estimated. It is known that the BIC implements a firmer penalty (Fitzmaurice et al., 2004). For both AIC and BIC, values closer to zero represent better fit (Fitzmaurice et al., 2004). Using an example longitudinal data set in medicine, Littell, Pendergast, and Natarajan (2000) illustrated how to model the covariance structures. They also examined the effects of choosing a covariance pattern on fixed-effects testing and on estimation/standard error of differences between treatment means.

3.2.8.1. Methods: Covariance Pattern Models

Covariance Pattern Models (CPM) can be understood as an extension of the RM MANOVA. Similar to MANOVA, for commonly used CPMs, timing of the repeated measures is fixed (i.e. subjects are measured at the same occasions). However, unlike MANOVA, incomplete data across the fixed time measurements are allowed.

Let us assume $i = 1, \cdots N$ is the index for individuals, and $j = 1, \cdots n_i$ represents the number of observations for subject i (notice that in MANOVA, $n_i = n$ for all subjects).

The model, written as a regression model in matrix form, is as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e_i}$$

where,

- $\boldsymbol{y}_i$ denotes the $n_i \times 1$ vector representing the response for subject i,
- $\boldsymbol{X}_i$ denotes the $n_i \times p$ matrix representing the values of the p predictors for subject i,
- $\boldsymbol{\beta}$ denotes the $p \times 1$ vector of regression parameters,
- $\boldsymbol{e_i}$ denotes the $n_i \times 1$ error vector for subject i,

Model assumptions – All of the assumptions can be summarized as $\boldsymbol{e_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, which implies $\boldsymbol{y_i} \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$.

Each individual matrix $\boldsymbol{\Sigma}_i$ can be treated as a submatrix of the overall $n \times n$ matrix $\boldsymbol{\Sigma}$. Therefore, if $n_i < n$, then the rows and columns in $\boldsymbol{\Sigma}$ corresponding to the missing time points for individual have been removed resulting in matrix $\boldsymbol{\Sigma}_i$.

In this section the assumption is that the *n* time points are equally spaced (see Núñez-Antón & Woodworth, 1994) on how this constraint can be relaxed).

The matrix $\boldsymbol{\Sigma}$**,** and each individual matrix $\boldsymbol{\Sigma}_i$, can be written as a function of a vector $\boldsymbol{\theta}$ of q parameters. Each of the forms that the variance-covariance matrix can take is expressed by a different number of parameters. The mathematical presentations of the most common structures for $\boldsymbol{\Sigma}$ are reviewed next.

Compound Symmetry is one of the simplest variance-covariance structures and assumes equal variances ($\sigma_1^2 + \sigma^2$) and covariances (σ_1^2) for the repeated measures. It only requires the estimation of two parameters (i.e. $q = 2$). The matrix form is written as the following:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 & \dots & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 & \dots & \sigma_1^2 & \sigma_1^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_1^2 & \sigma_1^2 & & & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{bmatrix}$$

As mentioned before, under this form, the variance of the outcome variable is the same, $\sigma_1^2 + \sigma^2$, at every occasion and the covariance between any two occasions is also the same, σ_1^2. While the advantage of this form is that it only requires two parameters, the assumption that measurements further away in time will have the same level of association as two consecutive measures is not very realistic.

First-Order Autoregressive Structure form ($AR(1)$), like CS, only depends on the estimation of two parameters (i.e. $q = 2$). However, this structure is more suitable for longitudinal data as it allows for the correlation between two measurement occasions to be a function of the lag between them. The covariance for time points j and j' is expressed as the following:

$$\sigma_{jj'} = \sigma^2 \rho^{|j-j'|}$$

where ρ denotes the first-order autoregressive parameter and σ^2 is the error variance.

The matrix representation of this structure can be written as:

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \ddots & \rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

As we can see, this form assumes homogeneous variance σ^2 across all repeated measures. It also assumes that the covariance between measurements of the same participant decays exponentially as a function of the time lags. Though an improvement over the CS, this structure is still restrictive since it assumes that the covariance between time points decreases according to an $AR(1)$. The next type of structure relaxes this condition.

Toeplitz or Banded Structure also represents diminishing correlations across time point lags, but is less restrictive than the $AR(1)$ since a specific parameter is assigned to each lag, namely $\sigma_{jj'} = \theta_k$, where $k = |j - j'| + 1$.

The matrix form can be written as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \dots & \theta_n \\ \theta_2 & \theta_1 & \theta_2 & \dots & \theta_{n-1} \\ \theta_3 & \theta_2 & \theta_1 & \ddots & \theta_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \theta_n & \theta_{n-1} & \theta_{n-2} & \dots & \theta_1 \end{bmatrix}$$

Note that this form still assumes homogeneous variance θ_1 across time points. However, unlike the $AR(1)$ form, the correlations between observations measured at different time points are not restricted and can take any value. In cases where n is large, this form also provides the freedom to set the correlations for the higher-order lags to zero to minimize the number of estimated parameters. Another disadvantage of the $AR(1)$ and Toeplitz forms is that they are not well suited for cases in which the time intervals are not the same or similar.

Unstructured Form, as its name indicates, is the least restrictive of all forms. The restriction that was present in the Toeplitz form, that variance be equal across time, is now lifted. This form is also suitable if time intervals are not similar. The matrix representing the unstructured form can be written as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \dots & \theta_{1n} \\ \theta_{21} & \theta_{22} & \theta_{23} & \dots & \theta_{2n} \\ \theta_{31} & \theta_{32} & \theta_{33} & \ddots & \theta_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \theta_{n1} & \theta_{n2} & \theta_{n3} & \cdots & \theta_{nn} \end{bmatrix}$$

Each covariance is given a specific parameter. Because $\boldsymbol{\Sigma}$, being a covariance matrix, is symmetric, there are $q = \frac{n(n+1)}{2}$ unique parameters to be estimated under this form. Notice that this is the same structure assumed by the MANOVA, but with the difference that under the CPM models incomplete data are allowed. Although this form represents a more reasonable representation of reality, it uses more degrees of freedom and therefore requires a larger sample size.

3.2.9. Review: Non-Linear Models

Considerable attention has been paid in the longitudinal data analysis literature to linear and GLMs for model fitting, estimation procedures, and hypothesis testing (Molenberghs & Verbeke, 2005; Verbeke & Molenberghs, 2000). Despite the usefulness and flexibility offered by the linear and generalized linear models, they are subject to constraints (Molenberghs & Verbeke, 2005). The term “linear model” represents functional forms that depend on parameters in a linear fashion (Molenberghs & Verbeke, 2005; Serroyen, Molenberghs, Verbeke, & Davidian, 2009). For GLMs, the formulations contain more complex non-linear dependence on parameters; parameters are included linearly at the level of a functions for predictors such as the Logit or Probit. However, parameters are transformed via a non-linear link function (for example, the Logit or Probit link) to describe the mean (Molenberghs & Verbeke, 2005; Serroyen et al., 2009).

While suitable transformations of the outcome and predictors can result in models that are appropriate in many practical scenarios, some data are essentially too non-linear in nature to be modeled linearly (e.g. using LMMs, GLMMs, and GEEs) (Molenberghs & Verbeke, 2005). Analyses of this kind concern a diverse spectrum of applied sciences such as pharmaceutical sciences (pharmacodynamics - pharmacokinetics), agriculture/forestry, manufacturing, education, and psychology, among others, where a common issue arising in data analysis is that the mechanisms governing the relation between the response variable and time cannot be modelled via the class of linear models.

In Educational and Psychology research, many researchers use research participants' (e.g.: students and patients) work, test scores, and progress to monitor skills development and to identify discrepancies in (academic) performance levels and trajectories between individuals over time. The research objective would be looking at the description of the functional form of growth (e.g.: linear or non-linear, increase or decrease, accelerate or decelerate) to model the correct functional form of growth (Nese, Lai, & Anderson, 2013). Growth curves depict the evolution of the quantity of a certain characteristic in time and are difficult to model linearly. Although it is possible to obtain an "approximate" fit by adopting a high-order polynomial approach or a generalization of linear models via (generalized) linear mixed-effect modeling or latent growth curve modeling (Blozis, Conger, & Harring, 2007; Molenberghs & Verbeke, 2005; Nese et al., 2013), an attempt like this would prove unsatisfactory, due to the non-linear nature of the observations. It is hence more appropriate to utilize a fully non-linear method, including but not limited to the Logistic growth curve model (Vonesh, 1992), the Gompertz model (Panik, 2014), the Weibull model (Panik, 2014), the power model (Panik, 2014), non-LMMs (Molenberghs & Verbeke, 2005; Serroyen et al., 2009), marginal non-linear models (Serroyen et al., 2009), and conditional non-linear models (Serroyen et al., 2009), and non-linear latent curve models (Blozis et al., 2007).

Although in linear and linear-mixed models, the parameter estimation must account for the correlation among measurement occasions (for the same subject), the interpretation is independent of the implemented covariance pattern (Serroyen et al., 2009). However, with non-linear models, various assumptions on the variability and correlation can lead to widely varying magnitudes and distinct interpretations for the regression parameters (P. Diggle, Heagerty, et al., 2002; Molenberghs & Verbeke, 2005). In addition, fitting non-linear models could be challenging due to the choice of starting values, convergence issues, and diagnostics (Molenberghs & Verbeke, 2005; Vonesh, 1992). Examples of non-linear modeling for repeated measure data in the Education and Psychology include Caprara et al. (2008); Cameron et al. (2014); Mok et al. (2015); and Dumas and McNeish (2017).

3.2.9.1. Methods: Non-Linear Models

Throughout this paper, we have seen examples of the three main approaches to analyzing longitudinal data: marginal (e.g. GEE), conditional (e.g. transitional models for categorical response or conditional LMMs for continuous response), and subject-specific or growth models (e.g. LMMs). The focus of *marginal* models is on the change in the marginal expectation or mean response over subpopulations that have the same values for X. In *conditional* models, the repeated measures are modeled conditionally on a subset of previous measurements. For *growth models*, the study of individual-level change across time is the focus.

All the models discussed so far have been considered in the setting of linear or GLMs. Note that the *non-linear model* is not the same as the GLM. Non-linear models refer to any function whose parameters are non-linear; in GLMs, by contrast, parameters are included linearly at the level of predictors after being transformed by link functions such as the Logit function for binary response data.

Most common applications of non-linear models are in pharmacokinetics and pharmacodynamics. These fields implement non-linear models due to the specific nature of the data they collect. As mentioned before, while in the linear and generalized linear context the parameter interpretation is the same regardless of the approach used (marginal, conditional, or random-effects), non-linear methods can lead to different interpretations of the regression parameters.

Although subject-specific or mixed-effects models are the approaches most broadly used for non-linear methods, in order to introduce these models one needs to revisit all three approaches using the following non-linear function of time (t) that approximates the 'S' shape commonly seen in growth data in nature,

$$\frac{\beta_1}{1 + e^{-\frac{t-\beta_2}{\beta_3}}}$$

In particular, this Logistic function was used to analyze growth curves of trunk circumferences in orange tree data (Serroyen et al., 2009). In the orange tree example, authors specify the parameters interpretations as follows:

- β_1 represents "the asymptotic circumference,"
- β_2 represents "the time at which half of the asymptotic value is reached,"
- β_3 represents "the curvature at the time half of the asymptotic value is reached."

A *non-linear random-effects model* for a response Y_{ij} at measurement *j* for subject i, is represented by:

$$E\left(Y_{ij} \middle| \boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}\right) = h(\boldsymbol{x}_{ij}, \beta, \boldsymbol{z}_{ij}, \boldsymbol{b}_i),$$

where,

- $\boldsymbol{b}_i$ represents the vector of random-effects,
- $\boldsymbol{x}_{ij}$ represents the vector of fixed effects variables,
- $\boldsymbol{z}_{ij}$ represents the variables corresponding to the random coefficients,
- β represents the vector of fixed-effects regression parameters,
- h represents a non-linear link function.

The following non-linear, mixed model is suggested to reflect the non-linear function in the example:

$$Y_{ij} = \frac{\beta_1 + b_i}{1+\exp[-\frac{t_{ij}-\beta_2}{\beta_3}]} + e_{ij},$$

$$b_i \sim N(0, \sigma_1^2),$$

$$e_{ij} \sim N(0, \sigma^2)$$

where b_i allows each 'intercept' $\beta_1 + b_i$ to vary across each subject (i.e. it allows the asymptotic circumference to vary across the different trees). While this approach is non-linear with respect to the fixed-effects, it is linear with respect to the random effects b_i.

This formula can be extended by adding the random effects, b_{i2} and/or b_{i3}, corresponding to the other two fixed-effects parameters. For example, the following formula also includes a random effect for the second parameter:

$$Y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp[-\frac{t_{ij} - \beta_2 - b_{i2}}{\beta_3}]} + e_{ij}$$

A *marginal non-linear model* for a response Y_{ij} at measurement j for subject i, is represented by

$$E\left(Y_{ij} \middle| \boldsymbol{x}_{ij}\right) = h(\boldsymbol{x}_{ij}, \beta),$$

where,

- x_{ij} represents a vector of covariates,
- β represents the regression parameter vector,
- h represents the non-linear link function.

Following with the initial example, let us assume a marginal model with serially correlated errors that follow a functional form $\exp(-\phi u)$, where u represents the lag between two measurements, $u_{jk} = |t_{ij} - t_{ik}|$. The resulting model is written as:

$$Y_{ij} = \frac{\beta_1}{1+\exp[-\frac{t_{ij}-\beta_2}{\beta_3}]} + e_{(1)ij} + e_{(2)ij},$$

where,

- The first error term $e_{(1)ij} \sim N(0, \sigma^2)$ is independent across subjects,
- The second error term $e_{(2)ij} \sim N(0, \tau^2 H_i)$,
    - With the elements of matrix H_i being $h_{i,jk} = \exp(-\phi u_{jk})$, represents the correlation between measurements of a same subject.

A *conditional non-linear model* will incorporate the subset of outcomes $\bar{Y}_{ij}$ as a component of *h:*

$$E(Y_{ij} | Y_{ik,k\neq j}, \boldsymbol{x}_{ij}) = h(\boldsymbol{x}_{ij}, \beta, \bar{Y}_{ij}, \alpha),$$

where α is the vector of parameters corresponding to both the autoregressive effects and the variance-component parameters. Further restricting the conditioning to previous observations warrants a complete specification for this model.

A conditional model for the orange tree data can be achieved by assuming a transition model where b_i in the previous formula for the non-linear random-effects model is now replaced by the prior observations. This model is written as:

$$Y_{ij} = \frac{\beta_1 + \gamma Y_{i,j-1}}{1 + \exp[-\frac{t_{ij} - \beta_2}{\beta_3}]} + e_{ij}$$

3.2.10. Review: Non-Parametric Linear Models

Recent challenges in analyzing complex longitudinal data have encouraged the evolution of more complex yet flexible methods for modeling repeated measure data. Non-parametric analysis approaches are more data-adaptive while less restrictive compared to parametric methods. Non-parametric approaches can offer a promising alternative for dealing with repeated measure data, though they are not the focus of this review. Researchers seeking more information on non-parametric approaches should see Ramsay and Silverman (2002) for an introduction to this topic.

4. Discussion

Change over time is an inherent property of data in Education and Psychology that is frequently examined in observational and/or experimental settings. This paper has offered a review of current practices regarding longitudinal models, an overview of the statistical details of each method, and an identification of the best methods available. All methods have been classified into traditional versus advanced models.

To come up with a comprehensive list of longitudinal analysis methods used in Education and Psychology, a survey of four journals was conducted. The survey conveyed that multilevel modeling approaches (i.e. HLM, LMM, and random-effect models) for longitudinal data are consistently among the top two most commonly used models in these disciplines. However, the covariance structures of repeated measure data implemented in this category of models are almost always not reported. This might mean scholars are using the default methods without considering the covariance structure of their repeated measure data. If this is the case, researchers may be choosing models that lack the precision in testing and estimation required for modeling longitudinal data.

The survey further showed that SEM-related approaches are commonly used; because this type of modeling falls into an entirely different category of model, it has not been discussed in detail in this paper. Methods such as GEE and CPM are barely used in the journals surveyed, even though (as shown in Table 7) GEEs have optimal properties and CPMs are an essential part of analyzing longitudinal data.

Traditional models such as ANOVA and simple linear regression also were well represented in the surveyed journals, even though software such as R and STATA are capable of running more advanced models and the required methodologies are detailed in many longitudinal data analysis books (P. Diggle, Diggle, et al., 2002; Fitzmaurice, Davidian, Verbeke, & Molenberghs, 2009; Singer & Willett, 2003; Verbeke et al., 2014). While these methods can sometimes deliver adequate properties for smaller data sets with simple covariance structures, about 91% of the studies surveyed had sample sizes over 100 (see Figures 1-4 and Table 3), meaning that advanced models would perform better in these cases. Furthermore, traditional methods, although easy to understand, implement rigid and unrealistic assumptions (e.g. CS) that are often not satisfied. A quick look at each methods section can clarify where the assumptions of each method tend to fall apart.

The survey revealed that most of the reviewed articles had two to five time points. Although choosing the number of time points could depend on the question of interest, studying the trend of an intervention with advanced methods requires a larger number of repeated measures. In terms of sample size, most of the reviewed papers had a sufficient number; however, sample sizes smaller than 30 do show up. To be able to utilize advanced models and implement covariance pattern modeling along with LMMs, a larger number of time points (i.e. 10 or more) with larger sample sizes (i.e. 150 or more) is preferable.

Table 7 presents a summary of all reviewed models and corresponding references used in this paper. Information summarized in Table 7 is intended to help researchers choose an appropriate model out of the most commonly used options. It can help scholars to identify the optimal methods relative to their data structure and the assumptions that can be made.

A quick look at the first and second columns of Table 7 reveals that most models cannot handle missing and unequally spaced data in a straightforward way. This problem can be alleviated by using a General Serial Covariance (GSC) model. A GSC model is an LMM with a random intercept that can incorporate the covariance structure in a continuous way (i.e. the use of spatial correlation structure). This allows the GSC model to handle missingness in a straightforward way. This model was first introduced by Diggle (1988). The GSC model is not listed in Table 7 as it was not one of the models found in the survey of Education and Psychology literature; however, this approach will be covered in detail in the next paper. The GSC model can implement time-varying covariates and is relatively robust to violation of assumptions and small sample sizes.

Although most models in Table 7 account for the covariance structure of the data, most covariance structures such as CS and Toeplitz have rigid and unrealistic assumptions. On the other hand, covariance structures such as UN are very flexible, but with a large number of time points, these covariance structures struggle with estimating too many parameters.

Table 7 can also help researchers to explore model characteristics such as the ability to account for the hierarchical structure of data, robustness to small sample size, and violation of assumptions. Other factors such as the type of outcome variables and whether time-varying covariates are allowable are listed as well.

Overall, looking at Table 7, LMM/HLM can be chosen as one of the most flexible models discussed here. The model assumptions of LMM/HLM are also relatively simple compared to models such as heterogeneity and mixture models. In the next paper, an extension of LMM (i.e. the GSC) will be explored using a simulation study that evaluates the testing and estimation properties of this model. The third paper is a tutorial paper that can guide researchers seeking to implement and interpret the GSC model using a real-world data set.

	Complete data needed?	Equally spaced measurements needed?	Accounts for covariance structure?	Robust to small sample size?	Time varying covariates allowed?	Accounts for hierarchical clustering?	Robust to violation of assumptions?	Type of data needed?
RM ANOVA (Hedeker & Gibbons, 2006)	Yes/No. Likewise deletion is implemented. Missing occasions are allowed, but the design assumes each subject is measured at the same occasion.	Yes, the model assumes that time points are fixed across subjects (i.e. subjects are measured at the same occasions).	Yes, but it assumes CS, which is a very rigid and unrealistic variance structure.	Not particularly robust if number of subjects is small (i.e. less than 20).	No	Yes/No, only a random effect is included for subject.	Not very robust. It is robust to Normality assumptions to some extent.	Continuous response variable.
MANOVA (Hedeker & Gibbons, 2006)	Yes, it requires data for all occasions on which subjects have been measured.	Yes, time points are assumed to be equally spaced and fixed.	Yes, a general form of the covariance matrix is assumed.	No	No	No	Not very robust.	Continuous response variable.
RM ANCOVA (Hedeker & Gibbons, 2006)	No	Yes, time points are assumed to be equally spaced and fixed.	Yes, but in a restricted form (same as RM ANOVA).	Not robust particularly if number of subjects is small (i.e. less than 20).	Yes	Yes, but in a restrictive form (i.e. only a random effect can be included for subject).	Not very robust.	Continuous response variable.
LMM/HLM (Hedeker & Gibbons, 2006; Rabe-Hesketh & Skrondal, 2008)	No, can only handle MCAR and MAR.	No, time points do not need to be the same across subjects and can be unequally spaced.	Yes, one *can* specify different covariance structure (e.g. the GSC model). However, a random-intercept model is just CS. A random intercept and slope model has unrealistic and complex covariance pattern implemented.	Generally, estimation is based on large-sample distribution of tests statistics. However, small sample inference such as Kenward-Roger correction exist (McNeish, 2017).	Yes	Yes, via including various random effects.	Yes, but depends on the type of model. For instance, a model with random coefficients is usually more robust than a random intercept-only model.	Continuous response variable.
GEE (Hedeker & Gibbons, 2006; Rabe-Hesketh & Skrondal, 2008)	No, can handle MCAR. Note that fixed number of measurements is needed.	Yes	Yes, through the working correlation matrix but it is treated as nuisance.	The small sample properties of the GEE estimates are unknown. Bias-correction estimates are needed (Paul & Zhang, 2014).	Yes	Yes/No, they can only account for a source of non-independence.	Pretty robust to mis-specification of the covariance structure but loses efficiency if mis-specified.	Any

Common Covariance Pattern Models (i.e. CS, AR(1), Toeplitz, and UN) (Hedeker & Gibbons, 2006)	No	Most commonly used variance structures assume equally spaced time points. Most importantly, time points are assumed to be fixed.	Yes, and the different structures listed on the left-hand column go from more to less rigid.	Not very robust. Small samples do not allow estimation of less rigid (more complex) variance structures.	Yes	No	Fairly robust.	Continuous response.
Heterogeneity Model (Verbeke & Molenberghs, 2000)	No	No (same as LMM).	Yes	Same as LMM.	Yes	Yes	Quite robust.	Any (only continuous model discussed here).
Transition Models as a specific conditional model. (Fitzmaurice & Molenberghs, 2009)	No, Missing observations are allowed but complicate calculations.	Yes/No, one can use different coefficients in case of unequally spaced data.	Yes, but it accounts for the covariance structure in a different way than using correlated errors/random effects.	Not very robust, as depends on number of switches between states (Erlandsson, 2005).	Yes	Yes, one can introduce random intercept.	Not enough research available.	Any
Mixture Models (Montfort, Oud, & Satorra, 2010)	No	No	Yes	Small sample corrections are proposed (Chretien, 2009).	Yes	Yes	Quite robust.	Any
Derived Variable Approach (Hedeker & Gibbons, 2006)	Yes/No, however missing data can be a problem since each observation carries different 'weight'. However, it mostly depends on the type of analysis used.	N/A	No	Yes/No, loss of power when collapsing data, hence more vulnerable to small sample size. However, can have higher power compared to more complex models.	No	No	Yes/No, depends on type of model used. Can be same as traditional regression models.	Any
Generalized LMM (Hedeker & Gibbons, 2006; Rabe-Hesketh & Skrondal, 2008)	No	No	Yes. One can define different var-cov pattern.	Yes, if variance-correction to small samples is applied.	Yes	Yes	Quite robust.	Any

Conditional LMM (Verbeke & Molenberghs, 2000)	No	No	Yes	Not well documented.	Yes	Yes	Robust to mis-specification of the cross-sectional covariates.	Continuous response.
Non-linear Models (Serroyen et al., 2009)	No	No	Yes	Depends on the approach used in the estimation.	Yes	Yes	Yes	Any
Non-parametric Models (Wang, 2014)	No	No	Yes, through defining non-parametric random effects.	No. Non-parametric methods require more data. However, they are more robust to small number of repeated measures.	Yes	Yes	Yes, as no distribution form is assumed.	Any

Table 7. *Summary of statistical method*

References

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, *102*(2), 281. https://doi.org/10.1037/a0019318

Aelterman, N., Vansteenkiste, M., Van Keer, H., & Haerens, L. (2016). Changing teachers' beliefs regarding autonomy support and structure: The role of experienced psychological need satisfaction in teacher training. *Psychology of Sport and Exercise*, *23*, 64–72. https://doi.org/10.1016/j.psychsport.2015.10.007

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allik, M., & Kearns, A. (2017). "There goes the fear": Feelings of safety at home and in the neighborhood: The role of personal, social, and service factors. *Journal of Community Psychology*, *45*(4), 543–563. https://doi.org/10.1002/jcop.21875

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, 93–114. https://doi.org/10.2307/271083

Ann M. Arthur, & Dawn L. Davis. (2016). A pilot study of the impact of double-dose robust vocabulary instruction on children's vocabulary grow. *Journal of Research on Educational Effectiveness*, *9*(2), 173-200. https:/doi.org/10.1080/19345747.2015.112687 5

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade english language learners. *Journal of Research on Educational Effectiveness*, *2*(4), 345–376. https://doi.org/10.1080/19345740903217623

Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology*, *107*(3), 805–820. https://doi.org/10.1037/edu0000008

Barnett, A. G., Koper, N., Dobson, A. J., Schmiegelow, F., & Manseau, M. (2010). Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology: Selecting the correct variance-covariance. *Methods in Ecology and Evolution*, *1*(1), 15–24. https://doi.org/10.1111/j.2041-210X.2009.00009.x

Bartl, H., Hagl, M., Kotoučová, M., Pfoh, G., & Rosner, R. (2018). Does prolonged grief treatment foster posttraumatic growth? Secondary results from a treatment study with long-term follow-up and mediation analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, *91*(1), 27–41. https://doi.org/10.1111/papt.12140

Bensley, D. A., Crowe, D. S., Bernhardt, P., Buckner, C., & Allman, A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in psychology. *Teaching of Psychology*, *37*(2), 91–96. https://doi.org/10.1080/00986281003626656

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An Introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*(2), 174–187. https://doi.org/10.1093/jpepsy/jst084

Berridge, D., Penn, R., & Ganjali, M. (2009). Changing attitudes to gender roles: A longitudinal analysis of ordina response data from the British Household Panel Study. *International Sociology*, *24*(3), 346–367. https://doi.org/10.1177/0268580909102912

Blonigen, D. M., Carlson, M. D., Hicks, B. M., Krueger, R. F., & Iacono, W. G. (2008). Stability and change in personality traits from late adolescence to early adulthood: A longitudinal twin study. *Journal of Personality*, *76*(2), 229–266. https://doi.org/10.1111/j.1467-6494.2007.00485.x

Blozis, S. A., Conger, K. J., & Harring, J. R. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development*, *31*(4), 340–346. https://doi.org/10.1177/0165025407077755

Boden, J. M., Van Stockum, S., Horwood, L. J., & Fergusson, D. M. (2016). Bullying victimization in adolescence and psychotic symptomatology in adulthood: Evidence from a 35-year study. *Psychological Medicine*, *46*(06), 1311–1320. https://doi.org/10.1017/S0033291715002962

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383. https://doi.org/10.1177/0049124103260222

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: a structural equation perspective*. In *Wiley Series in Probability and Statistics*. Hoboken, N.J: Wiley-Interscience.

Brailean, A., Aartsen, M. J., Muniz-Terrera, G., Prince, M., Prina, A. M., Comijs, H. C., … Beekman, A. (2017). Longitudinal associations between late-life depression dimensions and cognitive functioning: A cross-domain latent growth curve analysis. *Psychological Medicine*, *47*(04), 690–702. https://doi.org/10.1017/S003329171600297X

Breeman, L. D., Jaekel, J., Baumann, N., Bartmann, P., & Wolke, D. (2016). Attention problems in very preterm children from childhood to adulthood: The Bavarian Longitudinal Study. *Journal of Child Psychology and Psychiatry*, *57*(2), 132–140. https://doi.org/10.1111/jcpp.12456

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology*, *46*(2), 129–149. https://doi.org/10.1016/j.jsp. 2007.02.004

Brown, C. L., Gibbons, L. E., Kennison, R. F., Robitaille, A., Lindwall, M., Mitchell, M. B., … Piccinin, A. M. (2012). Social activity and cognitive functioning over time: A coordinated analysis of four longitudinal studies [Research article]. https://doi.org/10. 1155/2012/287438

Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Schilo, L., & Grissmer, D. W. (2014). Nonlinear Gompertz curve models of achievement gaps in mathematics and reading. *Journal of Educational Psychology*, *107*(3), 789. https://doi.org/10.1037/edu0000009

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, *100*(3), 525–534. https://doi.org/10.1037/0022-0663.100.3.525

Cemalcilar, Z., & Falbo, T. (2008). A longitudinal study of the adaptation of international students in the United States. *Journal of Cross-Cultural Psychology*, *39*(6), 799–804. https://doi.org/10.1177/0022022108323787

Chretien, S. (2009). Estimation of Gaussian mixtures in small sample studies using l1 penalization. *ArXiv:0901.4752 [Stat]*. Retrieved from http://arxiv.org/abs/0901.4752

Christens, B. D., & Speer, P. W. (2011). Contextual influences on participation in community organizing: A multilevel longitudinal study. *American Journal of Community Psychology*, *47*(3–4), 253–263. https://doi.org/10.1007/s10464-010-9393-y

Ciarrochi, J., Parker, P., Sahdra, B., Marshall, S., Jackson, C., Gloster, A. T., & Heaven, P. (2016). The development of compulsive internet use and mental health: A four-year study of adolescence. *Developmental Psychology*, *52*(2), 272–283. https://doi.org/10.1037/dev0000070

Cox, D. R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *21*(2), 113–120. https://doi.org/10.2307/2346482

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121–136. https://doi.org/10.1080/15248371003699969

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., … Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research*, *79*(1), 69–102. https://doi.org/10.3102/0034654308325581

DeGraff, J. V., DeGraff, N., & Romesburg, H. C. (2013). Literature searches with Google Scholar: Knowing what you are and are not getting. *GSA Today*, 44–45. https://doi.org/10.1130/GSAT175GW.1

Diggle, P., Diggle, P. J., Heagerty, P., Heagerty, P. J., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford, UK: Oxford University Press.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics*, *44*(4), 959–971. https://doi.org/10.2307/2531727

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work (Reading, Mass.)*, *20*(2), 159–165.

Dumas, D. G., & McNeish, D. M. (2017). Dynamic measurement modeling: Using nonlinear growth models to estimate student learning capacity. *Educational Researcher, 46*(6), 284–292. https://doi.org/10.3102/0013189X17725747

Edmunds, J. A., Unlu, F., Glennie, E., Bernstein, L., Fesler, L., Furey, J., & Arshavsky, N. (2017). Smoothing the transition to postsecondary education: The impact of the early college model. *Journal of Research on Educational Effectiveness*, *10*(2), 297–325. https://doi.org/10.1080/19345747.2016.1191574

Edwards, L. J. (2000). Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatric Pulmonology, 30*(4), 330–344.

Erlandsson, U. (2005). *Transition variables in the Markov-switching model: Some small sample properties*. (Working Papers, Department of Economics, Lund University; No. 25) Department of Economics, Lund University.

Facal, D., Guàrdia-Olmos, J., & Juncos-Rabadán, O. (2015). Diagnostic transitions in mild cognitive impairment by use of simple Markov models. *International Journal of Geriatric Psychiatry*, *30*(7), 669–676. https://doi.org/10.1002/gps.4197

Fitzmaurice, G. M., Davidian, M., Verbeke, G., & Molenberghs, G. (Eds.). (2009). *Longitudinal data analysis*. Boca Raton, LA: CRC Press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. In *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley-Interscience.

Fitzmaurice, G. M., & Molenberghs, G. (2009). Advances in longitudinal data analysis: An historical perspective. In G. M. Fitzmaurice, M. Davidian, G. Verbeke, & M. Molenberghs (Eds.), *Longitudinal data analysis*. Boca Raton: CRC Press.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal*, *21*(4), 413–436. https://doi.org/10.1007/s11145-007-9083-9

Garson, G. (2013). *Hierarchical linear modeling: Guide and applications*. https://doi.org/10.4135/9781483384450

Gibbons, R. D., Hedeker, D., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology*, *6*, 79–107. https://doi.org/10.1146/annurev.clinpsy.032408.153550

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association*, *56*(296), 841–868. https://doi.org/10.1080/01621459.1961.10482130

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823

Guo, Y., Sun, S., Breit-Smith, A., Morrison, F. J., & Connor, C. M. (2015). Behavioral engagement and reading achievement in elementary- school-age children: A longitudinal cross-lagged analysis. *Journal of Educational Psychology*, *107*(2), 332–347. https://doi.org/10.1037/a0037638

Gustafsson, J.-E. (2010). Longitudinal designs. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological Advances in Educational Effectiveness Research* (1st Edition). https://doi.org/10.4324/9780203851005-14

Han, S., Capraro, R., & Capraro, M. M. (2015). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement. *International Journal of Science and Mathematics Education*, *13*(5), 1089–1113. https://doi.org/10. 1007/s10763-014-9526-0

Hart, S. R., Musci, R. J., Slemrod, T., Flitsch, E., & Ialongo, N. (2018). A longitudinal, latent class growth analysis of the association of aggression and special education in an urban sample. *Contemporary School Psychology*, *22*(2), 135–147. https://doi.org/10.1007/s406 88-017-0160-z

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*(2), 81–102. https://doi.org/10.1177/1471082X0100100201

Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal data analysis*. In *Wiley Series in Probability and Statistics*. Hoboken, N.J: Wiley-Interscience.

Hellfeldt, K., Gill, P. E., & Johansson, B. (2018). Longitudinal analysis of links between bullying victimization and psychosomatic maladjustment in Swedish schoolchildren. *Journal of School Violence*, *17*(1), 86–98. https://doi.org/10.1080/15388220.2016.1222 498

Hermanto, N., & Zuroff, D. C. (2018). Experimentally enhancing self-compassion: Moderating effects of trait care-seeking and perceived stress. *The Journal of Positive Psychology*, *13*(6), 617–626. https://doi.org/10.1080/17439760.2017.1365162

Holden, J. E., Kelley, K., & Agarwal, R. (2008). Analyzing change: a primer on multilevel models with applications to nephrology. *American Journal of Nephrology*, *28*(5), 792–801. https://doi.org/10.1159/000131102

Hong, S., Yoo, S.-K., You, S., & Wu, C.-C. (2010). The reciprocal relationship between parental involvement and mathematics achievement: Autoregressive cross-lagged modeling. *The Journal of Experimental Education*, *78*(4), 419–439. https://doi.org/10.1080/00220970903292926

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Huynh, H., & S. Feldt, L. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational and Behavioral Statistics*, *1*, 69–82. https://doi.org/10.3102/10769986001001069

Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers and Education*, *56*(4), 1023–1031. https://doi.org/10.1016/j.compedu.2010.12.002

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. Retrieved from https://Otexts.org/fpp2/

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, *102*(3), 588–600. https://doi.org/10.1037/a0019682

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00727

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Kent, K. M., Pelham, W. E., Molina, B. S. G., Sibley, M. H., Waschbusch, D. A., Yu, J., … Karch, K. M. (2011). The academic experience of male high school students with ADHD. *Journal of Abnormal Child Psychology*, *39*(3), 451–462. https://doi.org/10.1007/s10802-010-9472-4

Ker, H. (2014). Application of hierarchical linear models/linear mixed-effects models in school effectiveness research. *Universal Journal of Educational Research, 2*(2), 173–180.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Simulation and Computation, 27*(3), 591–604. https://doi.org/10.1080/03610919808813497

Kim, P., Rigo, P., Leckman, J. F., Mayes, L. C., Cole, P. M., Feldman, R., & Swain, J. E. (2015). A prospective longitudinal study of perceived infant outcomes at 18–24 months: Neural and psychological correlates of parental thoughts and actions assessed during the first month postpartum. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01772

King, R. B. (2015). Sense of relatedness boosts engagement, achievement, and well-being: A latent growth model study. *Contemporary Educational Psychology*, *42*, 26–38. https://doi.org/10.1016/j.cedpsych.2015.04.002

Kinnunen, J. M., Lindfors, P., Rimpelä, A., Salmela-Aro, K., Rathmann, K., Perelman, J., … Lorant, V. (2016). Academic well-being and smoking among 14- to 17-year-old schoolchildren in six European cities. *Journal of Adolescence, 50*, 56–64. https://doi.org/10.1016/j.adolescence.2016.04.007

Kisa, Z., & Correnti, R. (2015). Examining implementation fidelity in America’s choice schools: A longitudinal analysis of changes in professional development associated with changes in teacher practice. *Educational Evaluation and Policy Analysis*, *37*(4), 437–457.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*, (4th ed.). New York, NY: Guilford.

Konishi, C., Hymel, S., Danbrook, M. C., & Wong, T. K. (2018). Changes in bullying in relation to friends, competitiveness, and self-worth. *Canadian Journal of School Psychology*, 0829573518765519.

Kwok, O. M., West, S., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research, 42*(3), 557–592.

Kwok, O.-M., Lai, M. H.-C., Tong, F., Lara-Alecio, R., Irby, B., Yoon, M., & Yeh, Y.-C. (2018). Analyzing complex longitudinal data in educational research: A demonstration with project

English Language and Literacy Acquisition (ELLA) data using xxM. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00790

Ladd, G. W., Ettekal, I., & Kochenderfer-Ladd, B. (2017). Peer victimization trajectories from kindergarten through high school: Differential pathways for children's school engagement and achievement? *Journal of Educational Psychology*, *109*(6), 826–841. https://doi.org/10.1037/edu0000177

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974.

Lee, J., & Zentall, S. (2015). Reading motivation and later reading achievement for students with reading disabilities and comparison groups (ADHD and typical): A 3-year longitudinal study. *Contemporary Educational Psychology*, *50*. https://doi.org/10.1016/j.cedpsych.2015.11.001

Lee, W. K., Milloy, M. J. S., Walsh, J., Nguyen, P., Wood, E., & Kerr, T. (2016). Psychosocial factors in adherence to antiretroviral therapy among HIV-positive people who use drugs. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *35*(3), 290–297. https://doi.org/10.1037/hea0000310

Li, Y., & Lerner, R. M. (2013). Interrelations of behavioral, emotional, and cognitive school engagement in high school students. *Journal of Youth and Adolescence*, *42*(1), 20–32. https://doi.org/10.1007/s10964-012-9857-5

Liang, K. Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual Review of Public Health*, *14*, 43–68. https://doi.org/10.1146/annurev.pu.14.050193.000355

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lininger, M., Spybrook, J., & Cheatham, C. C. (2015). Hierarchical linear model: Thinking outside the traditional repeated-measures analysis-of-variance box. *Journal of Athletic Training*, *50*(4), 438–441. https://doi.org/10.4085/1062-6050-49.5.09

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*(13), 1793–1819.

Little, T. D. (2013). *Longitudinal structural equation modeling*. In *Methodology in the Social Sciences*. New York: The Guilford Press.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. In *Advanced Quantitative Techniques in the Social Sciences*: *Vol. 7*. Thousand Oaks: Sage Publications.

Long, M. C. (2016). Generating hypotheses and upper-bound effect sizes using a large number of school characteristics and student outcomes. *Journal of Research on Educational Effectiveness*, *9*(1), 128–147. https://doi.org/10.1080/19345747.2015.107 0939

Lopes Cardozo, B., Gotway Crawford, C., Eriksson, C., Zhu, J., Sabin, M., Ager, A., … Simon, W. (2012). Psychological distress, depression, anxiety, and burnout among international humanitarian aid workers: A longitudinal study. *PLoS ONE*, *7*(9), e44948. https://doi.org/ 10.1371/journal.pone.0044948

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226. https://doi.org/10.11 46/annurev.psych.51.1.201

Mäkikangas, A., Bakker, A. B., Aunola, K., & Demerouti, E. (2010). Job resources and flow at work: Modelling the relationship via latent growth curve and mixture model methodology. *Journal of Occupational and Organizational Psychology*, *83*(3), 795–814. https://doi.org/10.1348/096317909X476333

Markowitz, A. J. (2018). Changes in school engagement as a function of No Child Left Behind: A comparative interrupted time series analysis. *American Educational Research Journal*, *55*(4), 721–760. https://doi.org/10.3102/0002831218755668

Martin, B. H., & Calvert, A. (2018). Socially empowered learning in the classroom: Effects of arts integration and social enterprise in schools. *Journal of Teaching and Learning*, *11*(2), 27–42. https://doi.org/10.22329/jtl.v11i2.5057

Martin-Martin, A., Orduna-Malea, E., Harzing, A.-W., & Delgado López-Cózar, E. (2017). Can we use Google Scholar to identify highly-cited documents? *Journal of Informetrics*, *11*, 152–163. https://doi.org/10.1016/j.joi.2016.11.008

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-Variate distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. https://doi.org/10.1214/aoms/ 1177731915

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. (1980). *Applied time series analysis for the social sciences*. Retrieved from https://trove.nla.gov.au/version/45215880

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. In *Wiley Series in Probability and Statistics. Applied Probability and Statistics Section*. New York, NY: Wiley.

McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research*, *52*(5), 661–670. https://doi.org/10.1080/00273171.2017.1344538

Menard, S. W. (2002). *Longitudinal research* (2nd ed). In *Sage University Papers Series, 07-76* (2nd ed). Thousand Oaks, CA: Sage.

Mok, M. M. C., McInerney, D. M., Zhu, J., & Or, A. (2015). Growth trajectories of mathematics achievement: Longitudinal tracking of student academic progress. *British Journal of Educational Psychology*, *85*(2), 154–171. https://doi.org/10.1111/bjep.12060

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. In *Springer Series in Statistics*. Retrieved from //www.springer.com/us/book/9780387251448

Monfort, S. S., Howe, G. W., Nettles, C. D., & Weihs, K. L. (2015). A longitudinal examination of re-employment quality on internalizing symptoms and job-search intentions. *Journal of Occupational Health Psychology*, *20*(1), 50–61. https://doi.org/10.1037/a0037753

Montfort, K. van, Oud, J. H. L., & Satorra, A. (Eds.). (2010). *Longitudinal research with latent variables*. Heidelberg, Germany: Springer Verlag.

Moskowitz, J. T., Carrico, A. W., Duncan, L. G., Cohn, M. A., Cheung, E. O., Batchelder, A., … Folkman, S. (2017). Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV. *Journal of Consulting and Clinical Psychology*, *85*(5), 409–423. https://doi.org/10.1037/ccp0000188

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*(4), 371–402. https://doi.org/10.1037/1082-989X.2.4.371

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463–469.

Myers, S. A. (2017). A longitudinal analysis of students' motives for communicating with their instructors. *Communication Education*, *66*(4), 467–473. https://doi.org/10.1080/03634523.2017.1313437

Neighbors, C., Lewis, M. A., Atkins, D. C., Jensen, M. M., Walter, T., Fossos, N., … Larimer, M. E. (2010). Efficacy of web-based personalized normative feedback: A two-year randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *78*(6), 898–911. https://doi.org/10.1037/a0020766

Nese, J. F. T., Lai, C.-F., & Anderson, D. (2013). *A primer on longitudinal data analysis in education. Technical Report #1320*. Retrieved from https://eric.ed.gov/?id=ED545257

Newsom, J. T. (2012). Basic longitudinal analysis approaches for continuous and categorical variables. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 143–179). New York, NY: Routledge.

Ng, T. W. H., Feldman, D. C., & Lam, S. S. K. (2010). Psychological contract breaches, organizational commitment, and innovation-related behaviors: A latent growth modeling approach. *The Journal of Applied Psychology*, *95*(4), 744–751. https://doi.org/10.1037/a0018804

Núñez-Antón, V., & Woodworth, G. G. (1994). Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors. *Biometrics*, *50*(2), 445–456.

Oxford, M. L., & Lee, J. O. (2011). The effect of family processes on school achievement as moderated by socioeconomic context. *Journal of School Psychology*, *49*(5), 597–612. https://doi.org/10.1016/j.jsp.2011.06.001

Panik, M. J. (2014). *Growth curve modeling: Theory and applications*. Hoboken, NJ: John Wiley & Sons.

Paul, S., & Zhang, X. (2014). Small sample GEE estimation of regression parameters for longitudinal data. *Statistics in Medicine*, *33*(22), 3869–3881. https://doi.org/10.1002/sim.6198

Pearson, E. S., & Hartley, H. O. (1976). *Biometrika tables for statisticians*. London, UK: Biometrika trust.

Pennings, H. J., Brekelmans, M., Sadler, P., Claessens, L. C., van der Want, A. C., & van Tartwijk, J. (2018). Interpersonal adaptation in teacher-student interaction. *Learning and Instruction*, *55*, 41–57.

Piasecki, T. M., Cooper, M. L., Wood, P. K., Sher, K. J., Shiffman, S., & Heath, A. C. (2014). Dispositional drinking motives: associations with appraised alcohol effects and alcohol consumption in an ecological momentary assessment investigation. *Psychological Assessment*, *26*(2), 363–369. https://doi.org/10.1037/a0035153

Piro, J. M., & Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychology of Music*, *37*(3), 325–347. https://doi.org/10.1177/0305735608097248

Pittman, L. D., & Richmond, A. (2008). University belonging, friendship quality, and psychological adjustment during the transition to college. *Journal of Experimental Education*, *76*(4), 343–361.

Pusponegoro, N. H., Notodiputro, K. A., & Sartono, B. (2017). Linear mixed model for analyzing longitudinal data: A simulation study of children growth differences. *Procedia Computer Science*, *116*, 284–291.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using stata* (2nd ed.). College Station, TX: Stata Press.

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. In *Springer Series in Statistics*. New York, NY: Springer.

Rapport, M. D., Alderson, R. M., Kofler, M. J., Sarver, D. E., Bolden, J., & Sims, V. (2008). Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): The contribution of central executive and subsystem processes. *Journal of Abnormal Child Psychology*, *36*(6), 825–837. https://doi.org/10.1007/s10802-008-9215-y

Rubin, M., Evans, O., & Wilkinson, R. (2016). A longitudinal study of the relations among university students' subjective social status, social contact with university friends, and mental health and well-being. *Journal of Social and Clinical Psychology*, *35*, 722–737. https://doi.org/10.1521/jscp.2016.35.9.722

Russell, B. S., Lee, J. O., Spieker, S., & Oxford, M. L. (2016). Parenting and preschool self-regulation as predictors of social emotional competence in 1st grade. *Journal of Research* *in Childhood Education*, *30*(2), 153–169. https://doi.org/10.1080/02568543.2016.1143414

Schonert-Reichl, K., & Lawlor, M. (2010). The effects of a mindfulness-based education program on pre- and early adolescents' well-being and social and emotional competence. *Mindfulness*, *1*, 137–151. https://doi.org/10.1007/s12671-010-0011-8

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. Retrieved from JSTOR.

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In *Handbook of Developmental Research Methods* (pp. 265–278). New York, NY: Guilford.

Serroyen, J., Molenberghs, G., Verbeke, G., & Davidian, M. (2009). Nonlinear models for longitudinal data. *The American Statistician*, *63*(4), 378–388.

Shelton, B. E., Hung, J.-L., & Baughman, S. (2016). Online graduate teacher education: Establishing an EKG for student success intervention. *Technology, Knowledge and Learning*, *21*(1), 21–32.

Shephard, K., Harraway, J., Jowett, T., Lovelock, B., Skeaff, S., Slooten, L., … Furnari, M. (2015). Longitudinal analysis of the environmental attitudes of university students. *Environmental Education Research*, *21*(6), 805–820. https://doi.org/10.1080/13504622. 2014.913126

Simons-Morton, B., & Chen, R. (2009). Peer and parent influences on school engagement among early adolescents. *Youth and Society*, *41*(1), 3–25.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sitzmann, T., & Ely, K. (2010). Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. *Journal of Applied Psychology*, *95*(1), 132.

Smith, J. D., Handler, L., & Nash, M. R. (2010). Therapeutic assessment for preadolescent boys with oppositional defiant disorder: A replicated single-case time-series design. *Psychological Assessment*, *22*(3), 593.

Sullivan, A. L., Kohli, N., Farnsworth, E. M., Sadeh, S., & Jones, L. (2017). Longitudinal models of reading achievement of students with learning disabilities and without disabilities. *School Psychology Quarterly*, *32*(3), 336.

Sun, Z., Rosen, O., & Sampson, A. R. (2007). Multivariate Bernoulli mixture models with application to postmortem tissue studies in schizophrenia. *Biometrics*, *63*(3), 901–909.

Tang, X., & Qu, A. (2016). Mixture modeling for longitudinal data. *Journal of Computational and Graphical Statistics*, *25*(4), 1117–1137.

Twisk, J. W. R. (2003). *Applied longitudinal data analysis for epidemiology: A practical guide*. Retrieved from https://trove.nla.gov.au/work/23261094

Uhls, Y. T., Michikyan, M., Morris, J., Garcia, D., Small, G. W., Zgourou, E., & Greenfield, P. M. (2014). Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. *Computers in Human Behavior*, *39*, 387–392.

Van Nguyen, H., Laohasiriwong, W., Saengsuwan, J., Thinkhamrop, B., & Wright, P. (2015). The relationships between the use of self-regulated learning strategies and depression among medical students: an accelerated prospective cohort study. *Psychology, Health and Medicine*, *20*(1), 59–70. https://doi.org/10.1080/13548506.2014.894640

Verbeke, G., Fieuws, S., Molenberghs, G., & Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, *23*(1), 42–59.

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, *91*(433), 217–221. https://doi.org/10.1080/01621459.1996.10476679

Verbeke, G., & Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *48*(3), 363–375. Retrieved from JSTOR.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. In *Springer Series in Statistics*. New York, NY: Springer.

Verbeke, G., Spiessens, B., & Lesaffre, E. (2001). Conditional linear mixed models. *The American Statistician*, *55*(1), 25–34. https://doi.org/10.1198/000313001300339905

Vermunt, J. K. (2010). Longitudinal Research Using Mixture Models. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 119–152). https://doi.org/10.1007/978-3-642-11760-2_4

Vonesh, E. F. (1992). Non-linear models for the analysis of longitudinal data. *Statistics in Medicine*, *11*(14–15), 1929–1954. https://doi.org/10.1002/sim.4780111413

Vos, N., van der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers and Education*, *56*(1), 127–137. https://doi.org/10.1016/j.compedu.2010.08.013

Wang, J.-L. (2014). Nonparametric regression analysis of longitudinal data. In *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat05552

Webb, T. L., Sheeran, P., & Luszczynska, A. (2009). Planning to break unwanted habits: Habit strength moderates implementation intention effects on behaviour change. *British Journal of Social Psychology*, *48*(3), 507–523.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics-Simulation and Computation*, *22*(4), 1079–1106.

Wolfinger, R. D. (1996). Heterogeneous variance: covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 205–230.

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69.

Zumbo, B. D. (1999). The simple difference score as an inherently poor measure of change: Some reality, much mythology. *Advances in Social Science Methodology*, *5*, 269–304.

Paper 2: A Simulation Study of Linear Mixed Modeling with Spatial Correlation for Longitudinal Data

Hedyeh Ahmadi

Teachers College, Columbia University

ABSTRACT

Paper 2: A Simulation Study of Linear Mixed Modeling with Spatial Correlation for Longitudinal Data

Hedyeh Ahmadi

Choosing the best covariance structure in the analysis of repeated measure data is essential for properly analyzing the data in hand. However, a survey of longitudinal publications in Education and Psychology in the previous paper showed that most scholars do not report the covariance structures used, which suggests that either researchers are exploring the covariance structure of the repeated measure but not reporting it or simply using software defaults. Furthermore, even though the data might be consistent with spatial covariance structure, researchers in Education and Psychology mainly use Hierarchical Linear Models (HLM) with only random intercept or HLM with random intercept/slope. This simulation study explored the effect of running these HLMs when the data are consistent with the General Serial Covariance models (GSC) with spatial covariance patterns (i.e. Exponential, Gaussian, and Linear). In addition, the effect of sample size and data type was explored in terms of modeling properties using three different types of simulated repeated measure data, namely, balanced discrete, unbalanced discrete, and unbalanced continuous with three types of spatial covariance patterns (i.e. Exponential, Gaussian, and Linear). A detailed comparison of the GSC model with spatial covariance patterns (i.e. Exponential and Gaussian) to two HLMs (i.e. random intercept only and random intercept/slope models) is presented in terms of estimation and testing properties, when the data are consistent with the GSC model with spatial covariance patterns. An examination of bias, standard error (SE), coverage probability, and power showed that, regardless of data type, the GSC model with either Gaussian or Exponential covariance structures yielded the best estimation (mostly in terms of SE since all the estimated

parameters were unbiased) and testing properties, when data are consistent with the GSC model with Exponential, Gaussian, and Linear covariance patterns. A HLM with random intercept/slope model can be labeled as the next best model, keeping in mind the relatively low power (although still in acceptable range even with sample size of 150) and the mathematical limitations of its covariance structure as derived in this paper. A random intercept-only model had a SE furthest from the “true” standard error and the coverage probabilities were consistently outside the confidence interval. Model convergence issues in R were also explored briefly.

Table of Contents

Introduction

Longitudinal studies are a type of research in which data are collected from the same individuals over time. Measuring the same individual over time results in non-independence of the collected measures. In practice, as categorized in the first paper, there exist two groups of models that can account for the correlation between the repeated measures. The first group of models are categorized as *traditional* approaches such as Repeated Measure ANOVA (RM ANOVA) and Multivariate ANOVA (MANOVA); these models are less flexible but can account for the mentioned correlation structure in a very rigid way. The second group of models such as Linear Mixed Modeling (LMM) and Generalized Estimating Equations (GEE) are categorized as *complex/advanced* and are very flexible; these modeling approaches can model the correlation structure in a more realistic way.

Modeling covariance structure is often a critical part of longitudinal data analysis. Accurate inference calls for appropriate correlation pattern modeling. There are many different ways to model the correlation pattern. One can use Hierarchical Linear Modeling (HLM) to account for this covariance structure, or model the covariance pattern itself. These two strategies can also be used together, which are called the General Serial Covariance (GSC) model. The GSC models are the focus of this simulation study.

Add-on methods have been developed to improve the statistical properties of the more advanced modeling approaches. For example, the GSC model can be thought of as an LMM where one can plug in different correlation structures. The common existing covariance structures such as Compound Symmetry (CS), Toeplitz, and Unstructured treat the covariance structure in a discrete way. However, each of these frequently used covariance structures has shortcomings. For example, CS, which is the simplest pattern, assumes that correlations are the same for each set of time lags, regardless of the length of each measurement interval. Toeplitz covariance does account for the time lag, but it assumes all time lags have their own correlation, and that this correlation is different for each time lag. One can say that Toeplitz is more realistic than CS, but increasing the number of parameters remains an issue—the same concern one has when using the Unstructured covariance pattern.

Furthermore, these covariance structures also cannot deal with missingness and irregularly spaced data in a straightforward way. On the other hand, spatial correlation structures (such as Exponential and Gaussian) model the covariance structure in a continuous way, thus missingness and irregularly spaced data are no longer problems. The spatial correlation structures also incorporate time lag into the covariance pattern, and only one parameter is estimated for the serial correlation pattern.

One can use HLM alone to account for the covariance pattern, but HLM with only random intercept induces a CS structure. On the other hand, HLM with random intercept and random slope induces an unrealistic covariance structure (derived in Appendix C) where the covariance is an increasing function of time lag.

The focus of this simulation study is on having continuous outcomes in which one can use LMM along with covariance structure modeling to account for within- and between-subject variation in repeated measure data analysis. This type of modeling, called the GSC model, was first introduced by Diggle (1988).

A survey of four journals in Education and Psychology in the first paper showed that although longitudinal research and multilevel modeling (i.e. HLM/LMM) are relatively common in these fields, it is not common to use multilevel models while also modeling the covariance structure to improve the model's statistical properties. The purpose of this paper is to introduce the GSC model to the fields of Psychology and Education by performing a simulation study to explore how the GSC model can improve testing and estimation properties over currently used methods. It will be assumed throughout this paper that the simulated data sets are consistent with the GSC model with spatial covariance structures (i.e. Gaussian, Exponential, and Linear covariance structure).

The simulation study section has two main purposes. First, it studies the *estimation* properties of the fixed effect by exploring bias and standard error (SE) of estimates. Second, it studies the *testing* properties by examining coverage probability of 95% confidence interval and the power of the Wald test to detect meaningful difference. The GSC models with spatial covariance patterns (i.e. Gaussian and Exponential) will be compared to HLM models with random intercept only and random intercept/slope. In addition, this paper also explores the effect of sample size and data type for these models.

Finally, while it may come as no surprise that GSC models with spatial covariance structures perform better when the data simulated are consistent with these types of models, it has already been shown in the first paper that scholars in Education and Psychology still use traditional methods or basic HLM models even when their data might exhibit a spatial correlation structure. This simulation study demonstrates just how poorly the basic HLMs perform when the data are consistent with the GSC model with a spatial covariance pattern. This paper therefore reaffirms the importance of checking the covariance structure before running any models (in the third paper in this dissertation, the variogram will be introduced as a tool for doing precisely this).

The General Serial Covariance Model

The GSC model can be seen as an LMM that can incorporate correlation pattern modeling. The GSC model can be specified as follows:

$$Y_{ij} = \mu_{ij} + \alpha_i + W_i(t_{ij}) + \varepsilon_{ij} \tag{1}$$

where $i = 1, ..., N$ is the subject index and $j = 1, ..., n_i$ is the measurement index. The GSC model assumptions and specifications can be listed as follows:

- The fixed part of the GSC model is embedded in the μ_{ij} where $\mu_{ij} = \mathbf{X_{ij}}^{\mathbf{T}}\beta$.
- The person-specific random intercept is defined as $\alpha_i \overset{iid}{\sim} N(0, \nu^2)$.
- The leftover error (also called measurement error) is defined as $\varepsilon_{ij} \overset{iid}{\sim} N(o, \sigma^2)$.
- The serial correlation within the repeated measure is defined as intrinsic stationary Gaussian process, $W_i(t_{ij})$, where,
  - $E(W_i(t_{ij})) = 0$
  - $cov(W_i(t_{ij}), W_i(t_{ik})) = \tau^2\rho(|t_{ij} - t_{ik}|) = \tau^2\rho(u)$ where u is the time lag between measurements for the same subject.
  - For example, one can specify $\rho(u) = e^{(-(\frac{u}{\phi})^c)}$ where $c = 1$ induces an Exponential serial correlation structure and $c = 2$ induces a Gaussian serial correlation structure. The rate of exponential decrease (sometimes called the range) is $\frac{1}{\phi}$. Note that for equally distanced measurements, an Exponential serial correlation is the same as AR(1); derivation of this equivalency is shown in Appendix A.
  - Another example would be a Linear serial correlation structure, which is defined as

$\rho(u) = 1 - \frac{u}{d}$ for $u < d$ and zero otherwise. The range for the Linear serial correlation structure is defined as d, after which the correlation is assumed to be zero.

  - Note that Exponential, Gaussian, and Linear correlation patterns are all called spatial correlation structures and the terminology was borrowed from spatial statistics.

- Thus, the GSC model has three sources of variation, namely, variation in the random intercept which comes from α_i; variation in the serial process which comes from $W_i(t_{ij})$; and variation in the measurement error which comes from ε_{ij}. In addition, it is often assumed that all these parameters are independent with:
  - $var(\alpha_i) = \nu^2$
  - $cov(W_i(t_{ij}), W_i(t_{ik})) = \tau^2 \rho(|t_{ij} - t_{ik}|)$
  - $var(\varepsilon_{ij}) = \sigma^2$
- $cov(Y_{ij}, Y_{ik}) = \nu^2 + \tau^2 \rho(u) + \sigma^2 I_{j=k}$
- $var(Y_{ij} - \mu_{ij}) = var(R_{ij}) = \nu^2 + \tau^2 + \sigma^2$

Note that an HLM random intercept-only model can be considered a GSC model without the serial correlation. Furthermore, an HLM random intercept/slope model is also a GSC model without the serial correlation, with the addition of a random slope. Using the serial correlation component in addition to a random intercept model (i.e. the GSC model) can improve some of the statistical properties of a random intercept HLM model. Furthermore, using the additional serial correlation is an alternative and more flexible way to account for the existing correlation pattern as compared to a random intercept/slope HLM model (which again can improve some of the statistical properties of a random intercept HLM model). In this paper, the former and latter claims will be explored using a simulation study.

Simulation Study Design

Literature Review

Electronic searches of simulation studies regarding the covariance structures in longitudinal data analysis (in particular, using the mixed-effects models and multilevel modeling) were conducted via Google Scholar (DeGraff, DeGraff, & Romesburg, 2013; Martin-Martin, Orduna-Malea, Harzing, & López-Cózar, 2017), using key terms such as longitudinal, covariance structure, simulation, mixed-effects models, and multilevel modeling. No date range or study fields were specified. Studies identified through online searches were then reviewed to determine whether they were simulation studies for covariance structures in longitudinal research. Four simulation studies were identified for studying the impact of mis-specifying the within-subject covariance structure in longitudinal data analysis (in particular, using the mixed-effects models and multilevel modeling), including Keselman, Algina, Kowalchuk, and Wolfinger (1998)(Field of Statistics); Kwok, West, and Green (2007) (Field of Educational Psychology); Barnett, Koper, Dobson, Schmiegelow, and Manseau (2010) (Field of Ecology); and Pusponegoro, Notodiputro, Sartono, et al. (2017) (Field of Statistics/Developmental Psychology).

Keselman et al. (1998) compared Akaike Information Criterion (AIC) and Baysian Information Criterion (BIC) to examine their effectiveness in detecting different covariance patterns for equal/unequal group sizes, and covariance matrices with unbalanced (across groups) in non-spherical repeated measure designs (with normal and non-normal data). They concluded that AIC and BIC were not effective in identifying the correct covariance pattern; on average, for all of the the 26 investigated distributions, the AIC criterion only chose the correct structure 47% of the time while the BIC resulted in the correct structure 35% of the time (Keselman et al., 1998).

The Monte Carlo study conducted by Kwok et al. (2007) explored the effect of mis-specifying the covariance structure in longitudinal data analysis under the multilevel modeling and mixed-modeling frameworks. Three types of misspecification were examined: (a) under-specification arises within nested covariance matrices when the true covariance matrix is more complex than the chosen covariance matrixs (b) over-specification arises within nested covariance matrices when the true covariance matrix is more constrained than the chosen covariance matrix; and (c) general mis-specification arises when the true covariance matrix and the chosen covariance matrix are not nested (Kwok et al., 2007). It was discovered that, with the multilevel model, under-specification and general-misspecification of the covariance pattern usually lead to overestimation of the variances of the random effects and standard errors of the growth parameter estimates, which resulted in lower statistical power for testing of the corresponding growth parameters (Kwok et al., 2007). An unstructured covariance pattern under the mixed-model framework usually resulted in underestimation of standard errors for the growth parameter estimates, which led to increased type I error for tests of the corresponding growth parameters(Kwok et al., 2007).

Using a simulation dataset for exploring effects of forest fragmentation on avian species richness over 15 years, Barnett et al. (2010) compared three methods for choosing the covariance pattern, namely, the AIC, the Quasi-Information Criterion (QIC), and the Deviance Information Criterion (DIC). The overall success rate for choosing the correct covariance structure was 80.6% for the AIC, 29.4% for the QIC and 81.6% for the DIC.

Pusponegoro et al. (2017) applied linear mixed-effects models and modeled different types of covariance structures (i.e. Unstructured (UN), Compound Symmetric (CS), Heterogeneous Compound Symmetric (CSH), First-order Autoregressive (AR(1)) and Heterogeneous First-order Autoregressive (ARH(1)) in a simulation study of children's growth differences based on different feeding methods. The selection criteria were the three fit indices: Negative 2-Residual Loglikelihood (-2RLL), AIC, and Schwarz's Bayesian Criterion (SBC). It was reported that the UN covariance pattern always produced the best fit, however considering the large number of parameters UN is very inefficient. On the other hand, authors reported that ARH(1) is a suitable alternative covariance pattern that is easier for interpretation purposes (Pusponegoro et al., 2017).

These simulation studies confirmed that failure to take account of the covariance among the repeated measures would result in incorrect estimates of standard errors of the parameters and could lead to misleading inferences. However, there was no consensus regarding which fit indices should be used for covariance structure selection and which covariance structure should be applied during the modeling process.

Finally, note that this simulation study does not use any model selection criteria, but instead evaluates the actual estimates using confidence bands or by simply comparing the estimates. This literature review did not identify any simulation study that evaluated the effect of data type, sample size, and covariance pattern of repeated measure data on the testing and estimation properties of the fixed-effect estimation (when data are consistent with the GSC model with Exponential, Gaussian, and Linear covariance structure); this simulation study therefore addresses this gap in the literature.

General Simulation Specifications

The description of simulations in this paper is divided into two sections: data simulation specifications and model specifications. These two sections will be presented separately; the reader is therefore cautioned not to confuse the correlation pattern implemented in the data simulation process with the correlation structure applied to the LMM in the modeling process.

Due to extended time needed for running all four models for each of the 27000 simulations, Columbia University's Habanero Shared High Performance Computing (HPC) Cluster was used with R software (R Core Team, 2018), version R.3.5.0. More details about HPC operating system and specifications can be found at https://cuit.columbia.edu/shared-research-computing-facility.

Using different HPC systems can affect the estimation, depending on the internal operating system and its specifications. Therefore, to ensure the results are replicable, the necessary information and the R code are presented in Appendix D. Although the causes for these slight differences observed in the results using personal computers versus HPC were investigated, the findings of this exploration are not included in this paper but are available upon request. This question offers an interesting subject of future research as scholars increasingly use HPC systems.

The R codes used for this simulation study along with seed specifications are included in Appendix D. Convergence issues raised from model fitting were resolved as follows:

- In using `lme()` function in `nlme` package in R, the lme control option `opt="optim"` was used in which, according to the R manual, uses the optimization method called `"L-BFGS-B,"` which was introduced by Byrd, Lu, Nocedal, and Zhu (1995). The method "allows box constraints, that is each variable can be given a lower and/or upper bound." Nocedal and Wright (1999) is a reference that can be used to learn more about this method. It is often impossible or difficult to know a priori which optimization function would work best for a specific data set. In this simulation study, the method `"L-BFGS-B"` provided a higher convergence rate.
- If all four models converged for a specific seed, that seed was recorded and used for the estimations; all four models needed to converge for the seed to be used. Otherwise, that seed

would be replaced by the next seed in the sequence.

- Note that the option of running a GSC model with a Linear covariance pattern was also explored. The results are not reported and are reserved for another paper due to high non-convergence issues and volatile behavior of the GSC model with a Linear covariance structure. Furthermore, because in Education and Psychology it is not realistic for the correlation within the repeated measure to go to zero, a GSC model with a Linear covariance structure would not be an appropriate choice in most cases.

One gray area that requires further consideration is whether there is a systematic bias when the non-convergence data are ignored. Due to relatively large sample sizes and the small number of problematic seeds, the fixed-effect parameter estimations were all virtually unbiased. For smaller sample sizes, however, it might be essential to find a way to check for this systematic bias. Ignoring the non-convergence data may induce bias for covariance parameters as well. This is an area that requires further study, and future research would benefit from an improved understanding of the topic, especially if testing for covariance parameters is of main interest to researchers. Because the main focus of this paper is to explore the testing and estimation properties of the fixed-effect component, these questions are reserved for another study.

Data Simulation Specifications

Overall, 1000 simulations for each of the 27 data settings were used, which is a three-way combination of the following specifications:

- Three different data types (which will be defined shortly) called Balanced Discrete, Unbalanced Discrete, and Unbalanced Continuous.
- Three different covariance structures, namely, Exponential, Gaussian, and Linear.

- Three different sample sizes, namely 150, 350, and 500.

The choice of sample size was based on the paper by Rochon (1991), which presents multiple tables with different specifications of repeated measure data and their corresponding sample sizes for two-group repeated measure experiments.

The data were simulated using the following GSC model:

$$\begin{aligned} Y_{ij} =& \mu_{ij} + \alpha_i + W_i(t_{ij}) + \varepsilon_{ij} \\ \mu_{ij} =& 10 + 0.5Time + 6Treatment + 2Time \times Treatment \end{aligned} \tag{2}$$

where all model specifications and assumptions remain the same as in the previous section. All the coefficients in the above equation were borrowed from Table 5.7 in Singer and Willett (2003) book with slight modifications.

Throughout this paper, *balanced data* is defined as observations taken at equal intervals where the number of observations are the same across individuals.*Unbalanced data* is defined as observations taken at unequal intervals and in which the number of observations are not the same across subjects.

Data and more specifically time have been simulated using three different structures as follows:

- *Balanced discrete data*, defined as observations made at equal intervals and at the same time for all individuals. For this data type, each individual has 15 repeated measures.
- *Unbalanced discrete data*, defined as observations still made at specific times but in which some individuals might have missing data. For this design, each individual has 10 to 15 repeated measures.

Variance Parameters	Implemented Values
ν	1.5
τ	2
σ	1
ϕ	3
d	3.5

Table 1. *Implemented variance parameters.*

- *Continuous data*, defined as observations not made at the same time and not at equal intervals. For this design, each individual has 10 to 15 repeated measures, and data are sporadically missing. However, since the data are simulated as continuous, an additional uniform distribution of (0, 0.25) is added to the original unequally spaced measurements. The reasoning behind choosing unif(0, 0.25) is that if, for example, a student missed more than $\frac{1}{4}$ of the test time, this student would have to take the next exam.

The specified variance components corresponding to the GSC model specification formula are shown in Table 1. The numerical values have been chosen according to our toy data explorations, using the Opposite Naming Score data set, first used by Willett (1988) and also presented in Chapter 7 of Singer and Willett (2003). Diggle (1988) was used for parameter estimations to adjust and to make an educated choice for all our covariance structure parameters. In choosing these variance components, a set of variances were chosen such that they were not too close to zero (so the likelihood function is not too flat) in order for the R program to be able to converge.

Model Specifications for Simulations

To summarize, there are three different types of data, namely balanced discrete, imbalanced discrete, and continuous. For each of these data structures, there are three different covariance structures, namely Exponential, Gaussian, and Linear. Then for each of these settings, sample sizes of 150, 350, and 500 are included. Overall, then, there are 27 data sets with 1000 simulations for each of them. Then for each of the 27000 simulated data sets, the following four models were fitted:

- The GSC model with Exponential covariance structure, here called CorEXP.
- The GSC model with Gaussian covariance structure, called CorGauss.
- HLM with only random intercept, called HLM1 (which will induce CS correlation structure).
- HLM with random intercept and random slope, called HLM2.

The focus of this simulation study is exploring spatial covariance patterns mainly because correlation structures such as CS, Toplitz, AR(1), and Unstructured are frequently used in Education, Psychology, and many other disciplines. By employing different data structures, the simulation explores the usefulness of the addition of spatial serial correlation patterns compared to the simplest form of HLM, which only implements a random intercept (i.e. CS), and a slightly more complex version of HLM that implements random intercept along with random slope of time.

In choosing the best covariance structure or HLM model, one can use criteria such as AIC, BIC, LR, or visualization methods such as variograms. While the use of these criteria is controversial, other tools such as variograms are informative for choosing the type of covariance structure to be modeled. However, for simulation purposes, variograms are not as practical because they would need to be visually inspected for all 27000 simulations. However, an inspection of randomly selected variograms showed that the visualization can distinguish between Exponential, Gaussian, and Linear covariance structures. The random intercept, measurement error, and serial correlation were also clearly showing up in the examined plots. Though this paper does not focus on model comparison using variogram visualization or numerical criteria such as AIC, BIC, and LR to choose the best model, the subject merits further research.

Metrics of Comparisons

Full numerical results of the simulations are reported in Tables B1 to B27 in Appendix B. Using Gelman, Pasarica, and Dodhia (2002) guidelines, these 27 tables were converted to 16 plots, each composed of a 3-by-3 matrix plot. Figures 1 to 16 show the results of the simulations; what follows are the guidelines on how to read and interpret each of the 16 plots:

- The first row corresponds to the balanced discrete *data*, the second row corresponds to the unbalanced discrete *data*, and the third row corresponds to the continuous *data* (see Simulation Specifications section for definitions).
- The first, second, and third columns correspond to *data* simulated with Exponential, Gaussian, and Linear covariance structures, respectively.
- Each plot contains four color-coded lines corresponding to four different *models*, defined in the previous section. One of these models is the "right" model; this means if, for example, the data have been simulated using an Exponential covariance structure, then the GSC model with Exponential covariance pattern is the right model and the rest of the models are misspecified models.
- Each 3-by-3 plot is the estimation for the following quantities for each coefficient of the intercept, time, treatment, and interaction:
  - Bias
  - Standard error (SE)

- Coverage probability
- Power

- The color-coded dotted lines in Bias plots are confidence intervals (i.e. $zero \pm 2 \times \frac{SE_{True}}{\sqrt{1000}}$) for each model; throughout this paper, these confidence intervals are called confidence bands for bias estimates. The horizontal black line is drawn at zero to help the reader navigate through the Bias plots.
- The dotted black line for the SE plots is drawn at one since the SE is presented as the ratio of the estimated SE to the "true SE" (i.e. the Monte Carlo SE).
- The middle dotted black line for the coverage probability plots are drawn at the 95% nominal value. The two lines above and below the nominal value line are the confidence intervals (i.e. $0.95 \pm 2 \times \sqrt{\frac{0.95\times(1-0.95)}{1000}}$) for each model; throughout this paper, these confidence intervals are called confidence bands for coverage probability estimates.

One of the most challenging steps in this simulation study was to extract the parameters from R output and then to match them to this paper's parametrization. There exist at least four different parametrizations of the spatial correlation structures and user caution is required in extracting the correct parameters. Guidelines shown in Martinussen, Skovgaard, and Sorensen (2012) were used to derive all of the equivalency formulas. The function shown in Appendix D will enable readers to extract parameters consistent with the parametrizations shown in this paper.

For calculating coverage probability, the simulation uses a 95% confidence interval with Normal distribution. The same confidence interval was used to calculate the power of a two-sided Wald test for each coefficient. The β values under the null and alternative hypotheses were defined as follows:

$$\beta_{H_0} = (\beta_0, \beta_1, \beta_2, \beta_3) = (10, 0.5, 6, 2)$$

$$\beta_{H_1} = (\beta_0, \beta_1, \beta_2, \beta_3) = (11, 0.6, 7, 2.1)$$

The logic behind choosing the null hypothesis values is explained in the Data Simulation Specifications section above. β_{H_1} were used to simulate data under the alternative hypothesis for the power calculations. The alternative hypothesis values, which are related to the magnitude of the effect sizes, were chosen using the following rationale:

- The intercept coefficient under the null hypothesis was 10. An effect size of 1 was chosen for this coefficient, thus the alternative hypothesis was 11. To be conservative, $\frac{1}{10}$ of the null hypothesis was chosen as an effect size. Note that choosing an effect size based on standardization was not possible since the design matrix contains a column of all ones for the intercept.
- The effect size for the time coefficient was chosen based on standardization of the coefficient where one can detect a standardized effect size of 0.5, which is considered a medium standardized effect size. For illustration purposes, a short proof on how to choose an effect size using standardization for a continuous variable in the simplest case is provided below:

$$Y_i = \beta_0 + \beta_1 X_i$$

$$Y_i = \widehat{\beta_0} + \widehat{\beta_1}\frac{(X_i - \mu)}{\sigma} = (\widehat{\beta_0} - \widehat{\beta_1}\frac{\mu}{\sigma}) + \frac{\widehat{\beta_1}}{\sigma}X_i$$

$$\implies \beta_1 \sim \frac{\widehat{\beta_1}}{\sigma} \implies \widehat{\beta_1} \sim \beta_1\sigma$$

Hence, for example, if the standardized effect size of 0.5 (i.e. a medium standardized effect size) is of practical interest, then one can complete the calculations as follows, where σ is the standard deviation of the covariate X_i:

$$\widehat{\beta_1}_{H_1} - \widehat{\beta_1}_{H_0} = 0.5$$

$$\implies \sigma\beta_{1H_1} - \sigma\beta_{1H_0} = 0.5$$

$$\implies \beta_{1H_1} = \frac{0.5}{\sigma} + \beta_{1H_0}$$

The last line of the above derivation was used to come up with the value 0.1 as an effect size to get to 0.6 for the alternative hypothesis value of β_1.

- A similar process to the above derivation was used to calculate the effect size to be added to the treatment and the interaction coefficient (in order to detect the standardized effect size of 0.5).

What follows is the detailed interpretation of the simulation results shown in Figures 1 to 16.

Simulation Results

The simulation results will be evaluated based on estimation (i.e. bias and SE) and testing properties (i.e. coverage probability and power). Annotated R codes corresponding to all of the simulations are shown in Appendix D.

Before presenting the results, it is worth mentioning a few general statistical details:

- It is known that when using LMM, the fixed-effect estimates are mainly unbiased but, for completeness, Bias plots are presented along with the relative confidence intervals.

- The effect of an under-powered study can be very problematic because the test might not pick up on a meaningful effect (i.e. type I error). The effect of Type I error will be more pronounced with small sample sizes, which will not be an issue for this simulation study since all of our sample sizes are relatively large.
- On the other hand, having an over-powered study might means finding significant results for negligible effect sizes that are not practically important. However, this should not be an issue if one is careful about the magnitude of a meaningful effect size. Additionally, if working with human or animal subjects, an over-powered study with a large sample size might raise ethical concerns. Otherwise, for a cautious researcher who interprets the results while keeping statistical significance separate from practical significance, having an over-powered study does not present a statistical problem.

Intercept Simulation Results

Intercept: Bias. In general, it is understood that fixed-effect estimations are unbiased and Figure 1 confirms this theory. Looking at the range of the y-axes for all of the plots, one can observe that all of the bias estimates are from -0.01 to 0.01, which is within the estimated confidence bands. Regardless of whether the model is right or wrong, the fixed-effect estimates of intercepts are unbiased, as expected. Note that one would expect that as sample size increases, the bias decreases. However, this pattern is not consistently observed. One explanation could be that all of the estimates are unbiased (i.e. very close to the horizontal dotted line at zero) and the magnitude of bias is so small that the effect of sample size is not as prominent. Also, note that the magnitude of the y-axes are so small that, in reality, this observed increase or decrease is not as pronounced as shown in Figure 1. Finally, all of the lines stand close to one another with similar patterns (in each plot) for all data types, so in terms of bias for intercept all models are performing equally well, regardless of data type.

Intercept: Standard Error Ratio. Figure 2 shows the SE ratio (i.e.$\frac{EstimatedSE}{TrueSE}$) for the intercept with different data structures. Regardless of data type and sample size, one can observe that HLM1 consistently has an SE ratio of less than one. This means that HLM1's estimated SE is smaller than the true SE for the intercept coefficient. The SE ratios of the three other models are generally close to one, regardless of sample size and data type. This means that all of the models except HLM1 are very close to the true SE.

Intercept: Coverage Probability. As Figure 3 shows, HLM1 consistently has the lowest coverage probability, which is very much outside the confidence band. In all of the plots (except for the continuous and balanced Gaussian data), Gaussian/Exponential GSC and HLM2 models follow the same pattern and they are all very close to or above the nominal value of 95%. Thus, regardless of data type and sample size, coverage probability for the intercept coefficient is almost always close to the nominal value (and within the confidence band) for HLM2, Exponential, and Gaussian GSC models.

Intercept: Power. As Figure 4 shows, the pattern of the power for all four models follows the same trajectory: the power plateaus toward 100% power between sample sizes of 350 and 500. Note that with sample size of 150, HLM1 consistently performs the best, regardless of data covariance structure, however because HLM1 had the lowest coverage probability it is not fair to compare its power with the other models here. Ignoring the HLM1 model, the Gaussian models have the highest power for all data types by up to 9%. On the other hand, HLM2 consistently has the lowest power by up to 9%. However, all models have relatively high power for the intercept estimates. Thus, in general, all models perform very well in terms of power for all sample sizes regardless of data type.

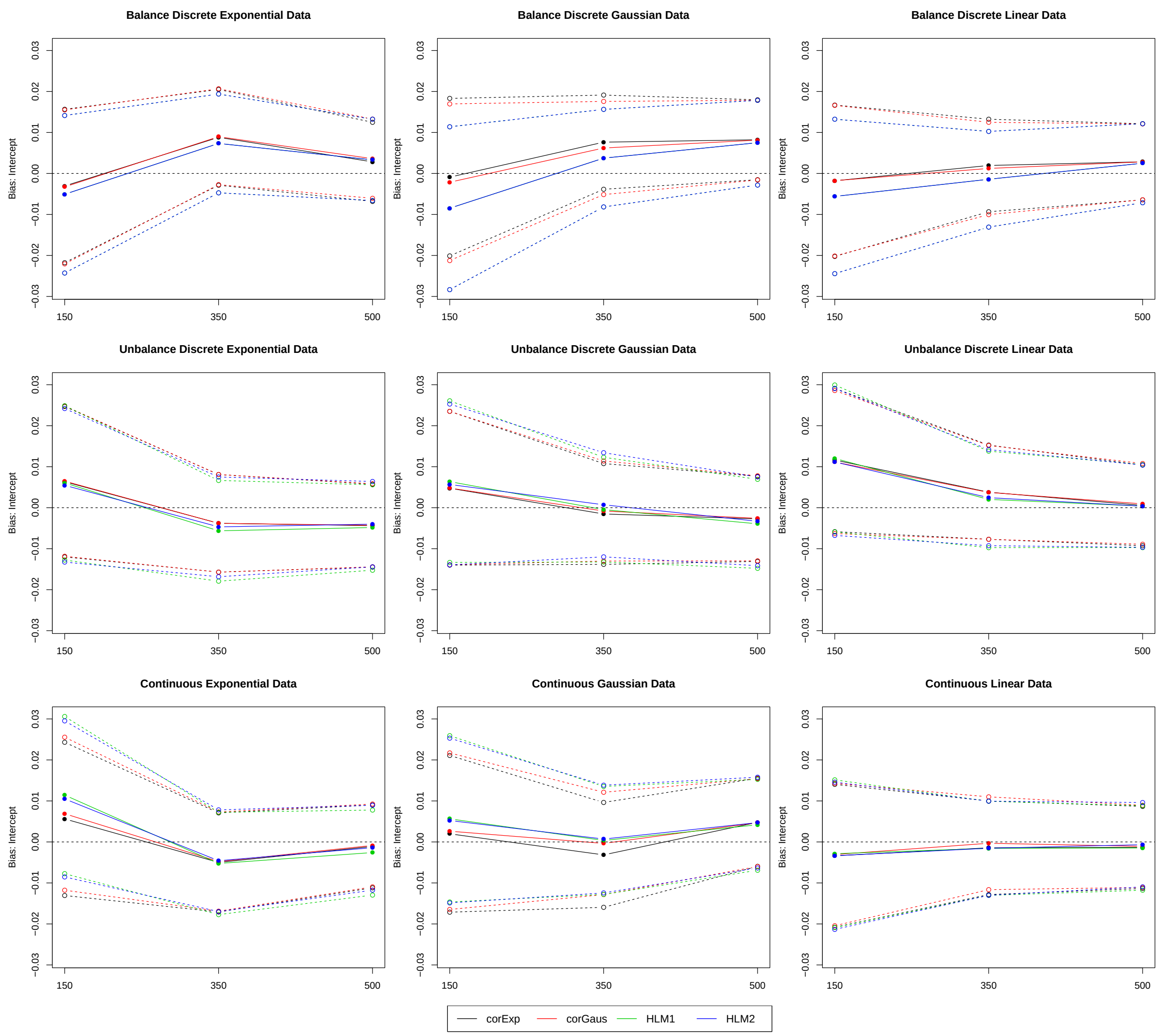


Figure 1. *Bias for intercept with different data specifications*

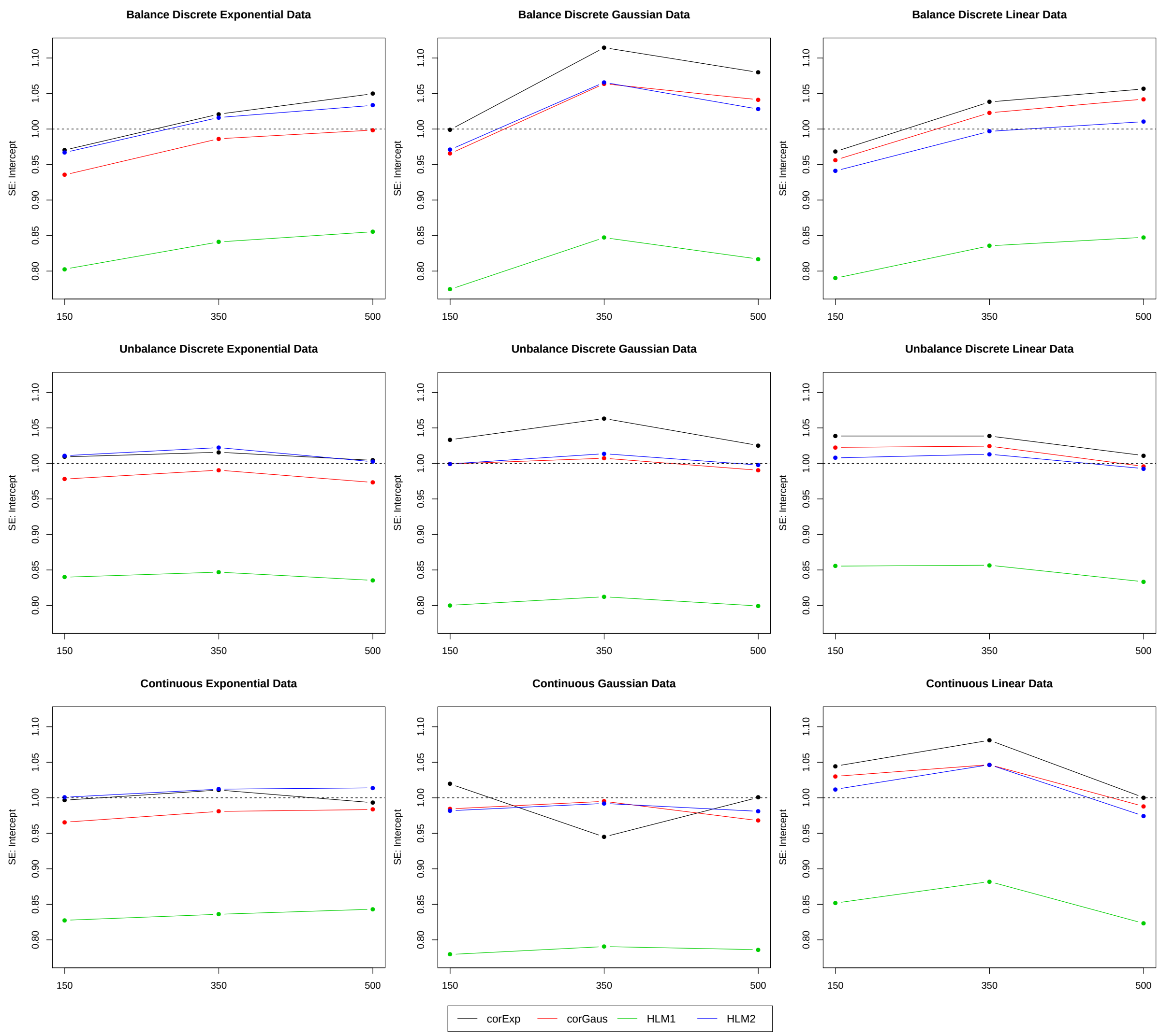


Figure 2. *Standard error ratio for intercept with different data specifications*

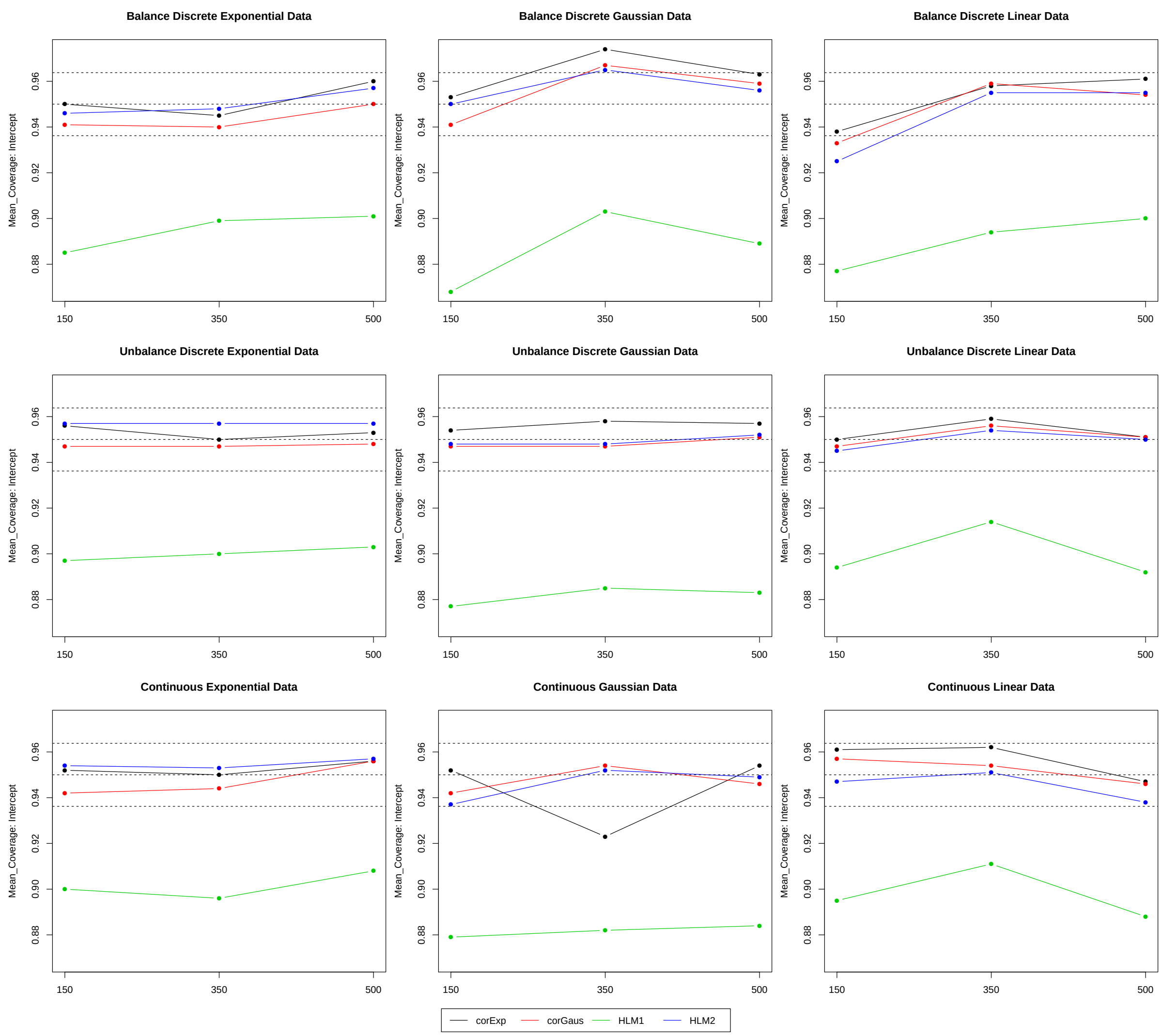


Figure 3. *Mean coverage for intercept with different data specifications*

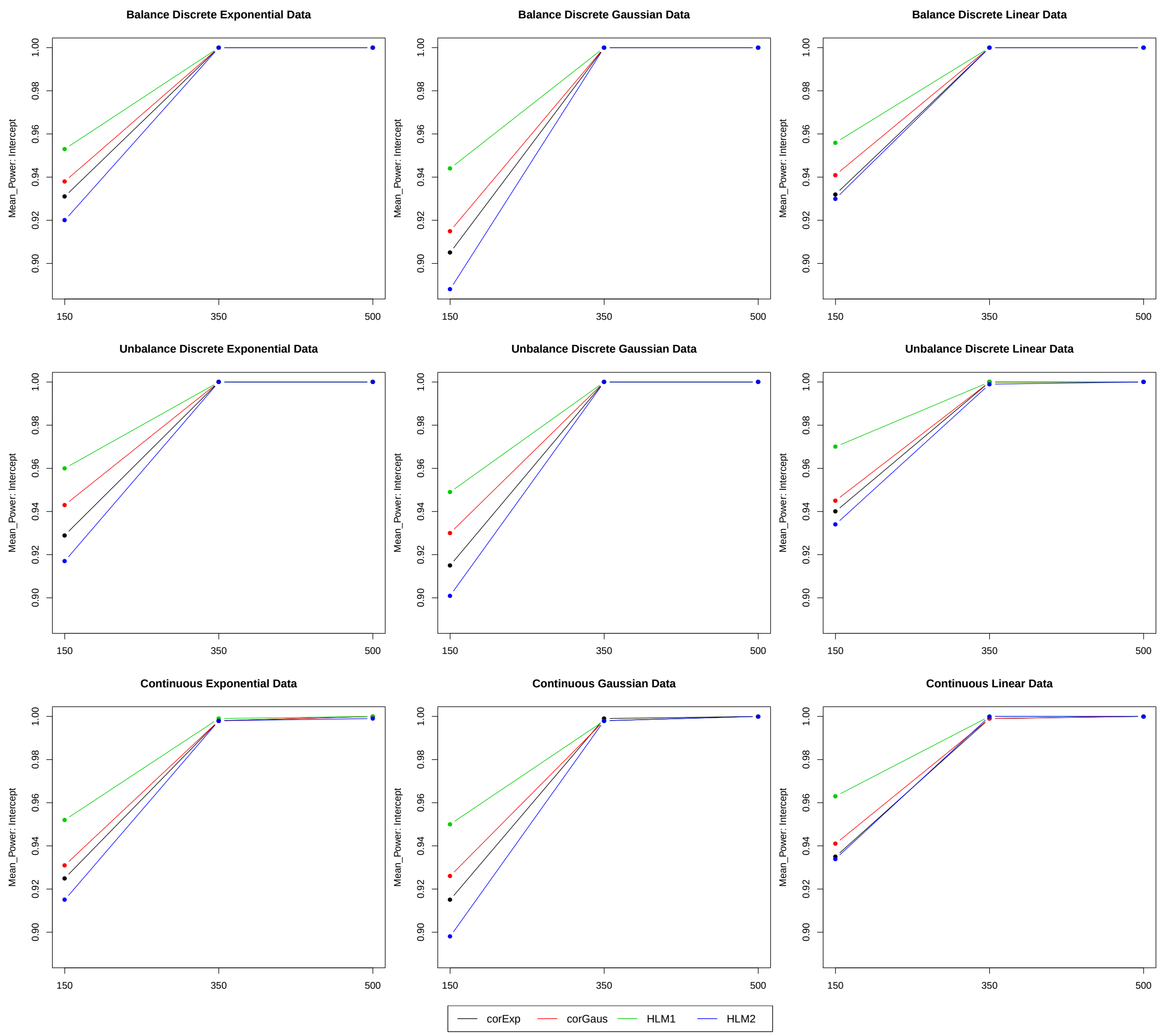


Figure 4. *Mean power for intercept with different data specifications*

Time Coefficient Simulation Results

Time Coefficient: Bias. Figure 5 shows that all the estimations for the time coefficient are unbiased. Note that, despite the volatility of some of the plots, all the y-axes are from -0.003 to 0.002. Thus, in terms of time coefficient, all estimates are virtually unbiased and within the presented confidence bands, regardless of data type, sample size, and choice of modeling.

Time Coefficient: Standard Error. Figure 6 shows that HLM1 consistently has the lowest SE ratio below one, regardless of data type and sample size; this means that again HLM1's estimated SE is smaller than the true SE for the Time coefficient. Overall, other than HLM1, all models' SE ratio estimates are close to one, which means in terms of SE, all the estimates are close to the "true" SE, regardless of data type and sample size. However, as expected, it is noticeable that when the data generating process matches the modeling technique the SEs are closer to the "true" SE. For the GSC with Linear covariance structure data (the last column of plots), all models except for HLM1 are following nearly the same trajectory.

Time Coefficient: Coverage Probability. Figure 7 shows that HLM1 consistently has the lowest coverage, making it the worst model for time coefficient in terms of coverage probability. As expected, when the data generating process matches the choice of modeling, the estimation of coverage probability falls very close to the nominal value of 95% and within the shown confidence band. Furthermore, for the GSC Exponential data, the GSC Gaussian model has the lowest coverage probability, which is sometimes slightly outside the confidence band for Time coefficient. For the GSC Linear data, the GSC models and HLM2 have acceptable coverage probabilities. Overall, in terms of Time coefficient, Exponential GSC, Gaussian GSC, and especially HLM2 are all acceptable models for coverage probability, regardless of data type and sample size.

Time Coefficient: Power. Figure 8 shows that for all data types, the power plateaus at sample sizes of 350 and 500. Again, HLM1 will be excluded from our discussion here since its coverage probability was not in an acceptable range. The Gaussian GSC consistently has the highest power and HLM2 consistently has the lowest power for sample sizes of 150; the difference between the former and the latter model can be up to 5%. In general, both GSC models and the HLM2 model have high mean power for the Time coefficient, regardless of data type and sample size (with HLM2 having the lowest among the three models).

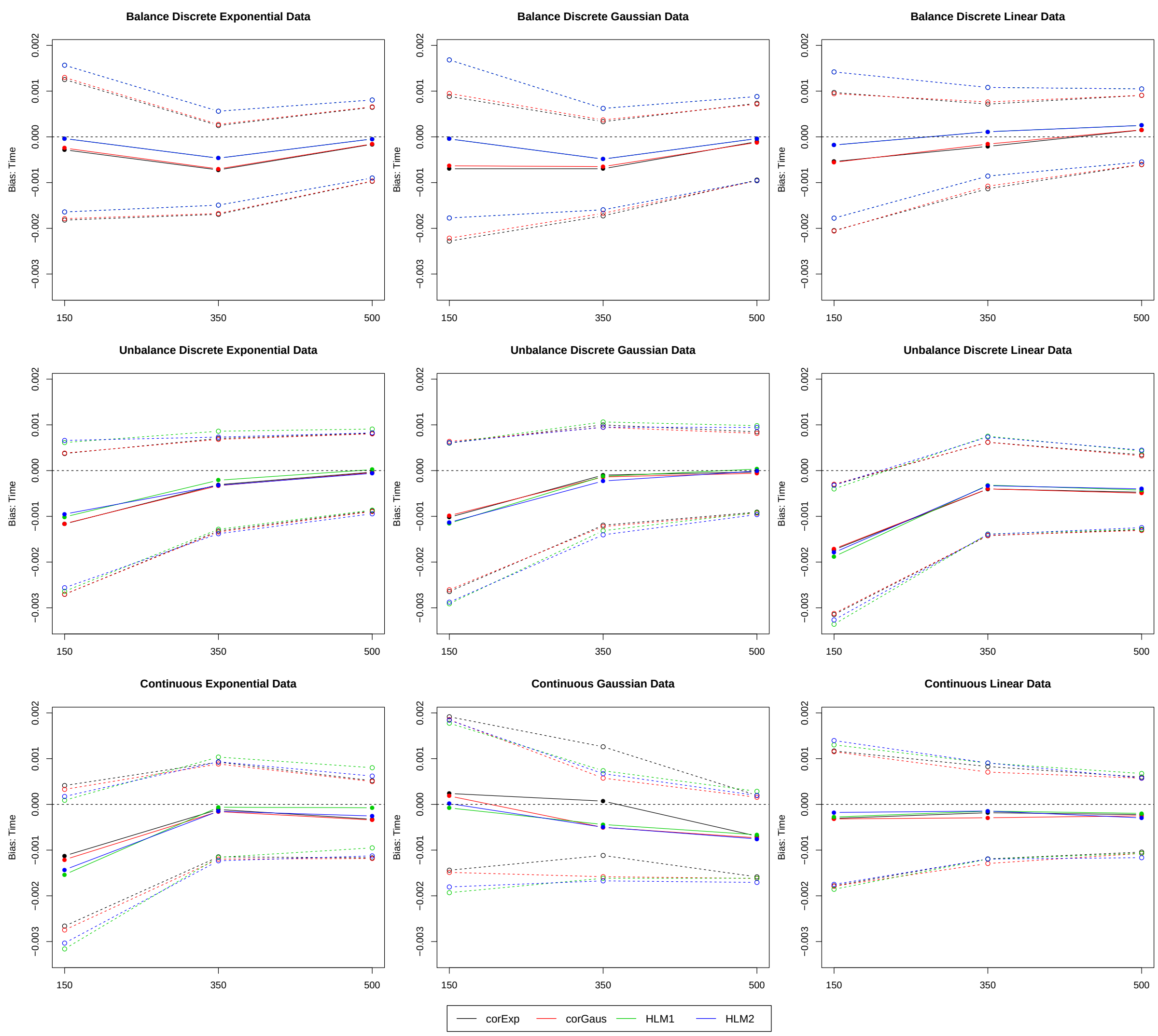


Figure 5. *Bias for the time coefficient with different data specifications*

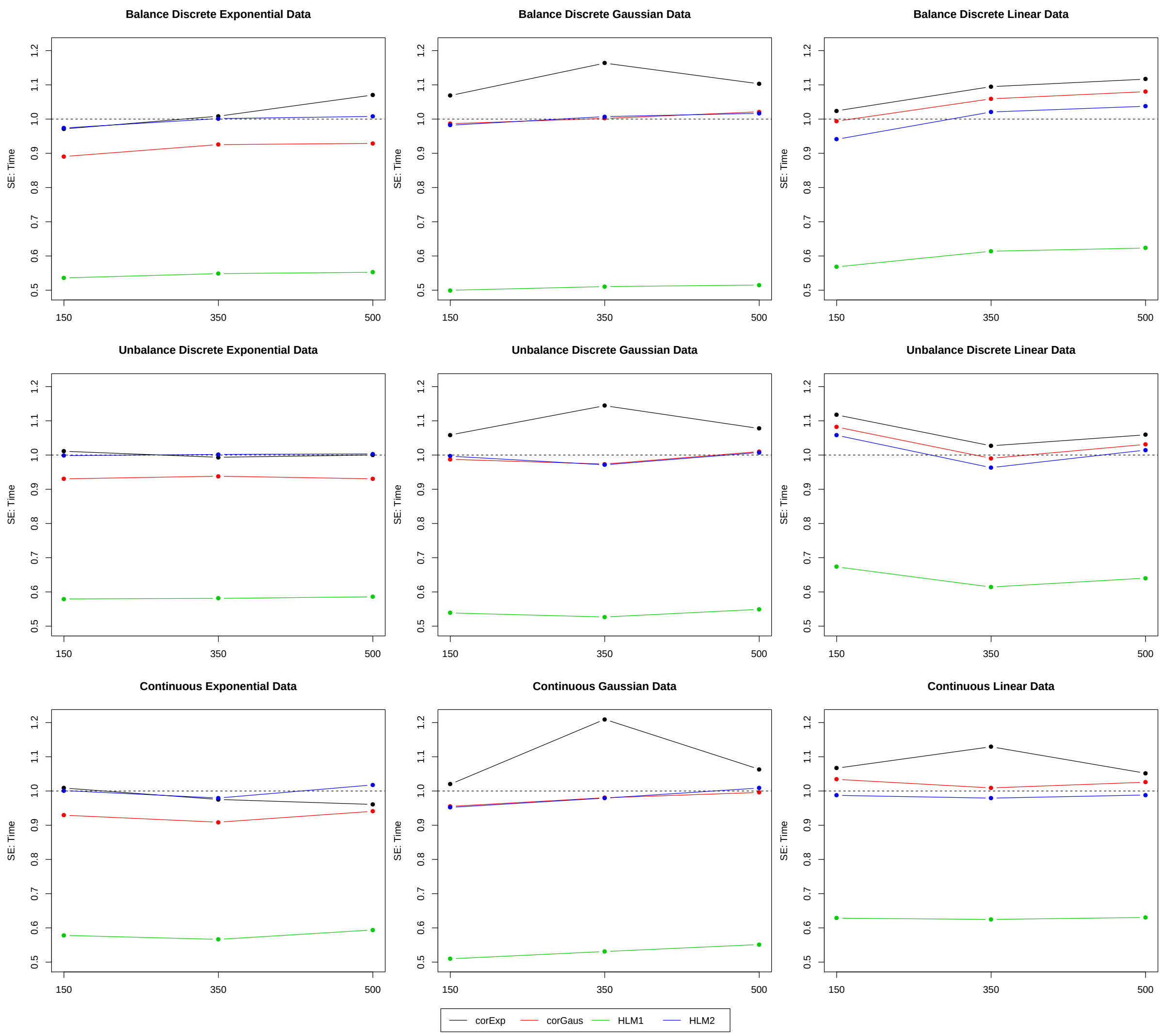


Figure 6. *Standard error for the time coefficient with different data specifications*

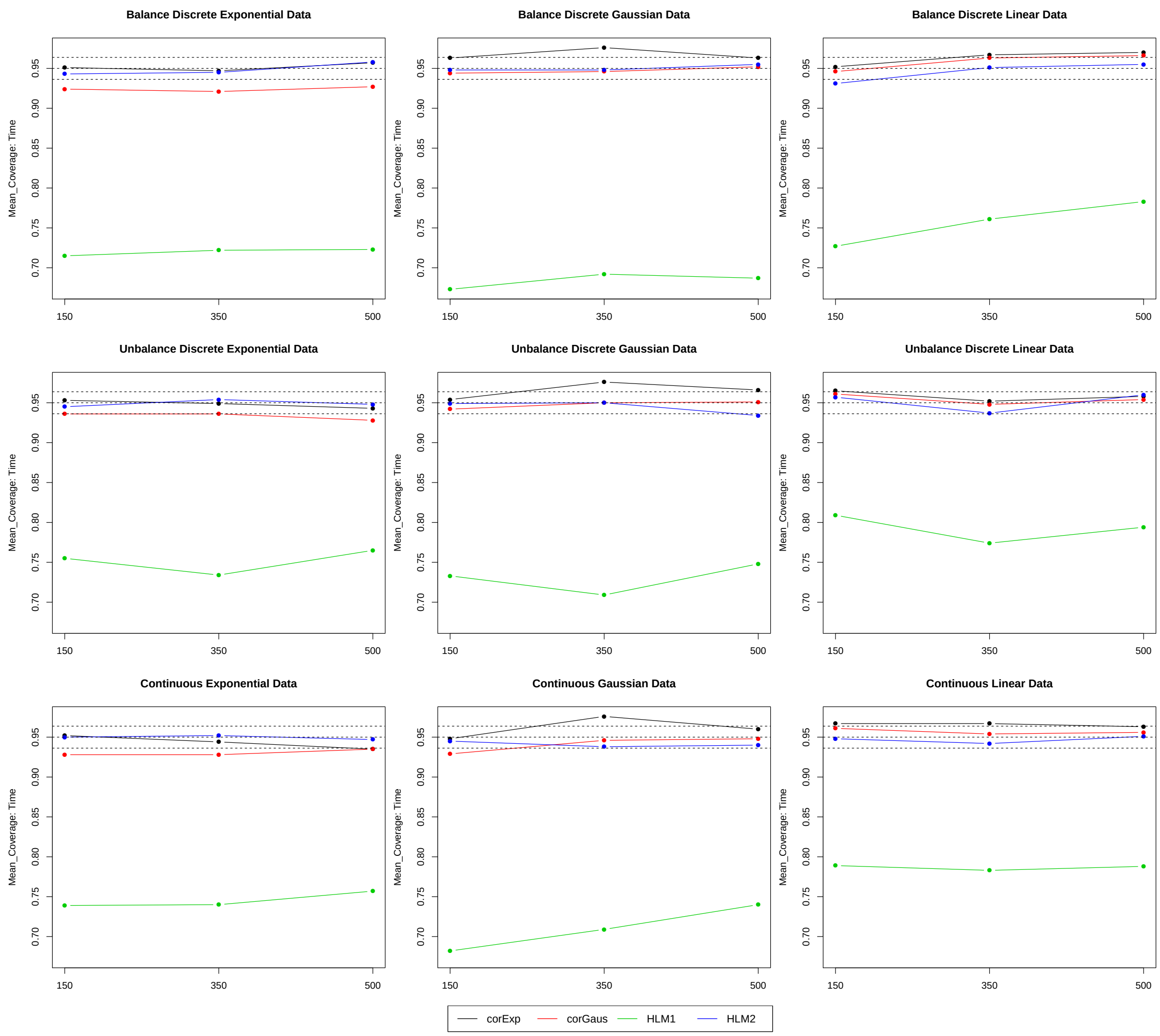


Figure 7. *Mean coverage for the time coefficient with different data specifications*

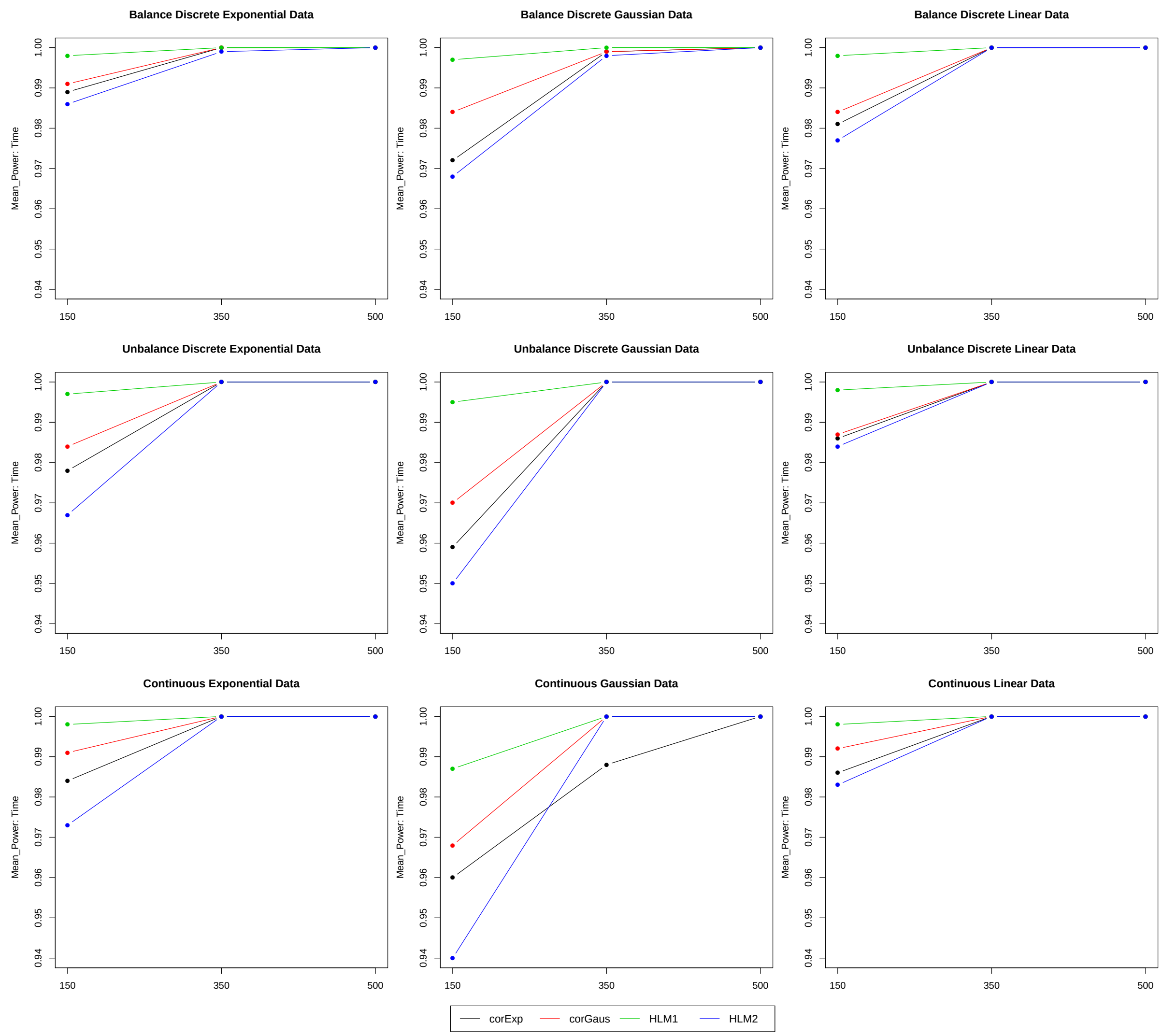


Figure 8. *Mean power for the time coefficient with different data specifications*

Treatment Coefficient Simulation Results

Treatment Coefficient: Bias. Figure 9 shows that all of the estimates for treatment coefficient have very small to negligible bias and thus are within the shown confidence bands. Although one would expect bias to decrease as sample size increases, this is not true for some of our plots; the magnitude of the y-axes is so small that the observed abnormal behavior is not as pronounced as it looks. In general, all estimates are unbiased, regardless of data type, sample size, and modeling choice.

Treatment Coefficient: Standard Error. Figure 10 shows that HLM1 consistently has the smallest SE ratio (less than one), regardless of data type and sample size. This means that again HLM1's estimated SE is smaller than the true SE for the Time coefficient. Overall, the three models other than HLM1 have SE ratio estimates that are close to one, which means that all the estimates are close to the "true" SE, regardless of data type and sample size. However, as expected, it is noticeable that when the data generating process matches the modeling technique the estimated SEs are closer to the "true" SE. For the GSC with Linear covariance structure data (the last column of plots), all models excepting HLM1 are following nearly the same trajectory.Finally, note that for all data types, Exponential GSC, Gaussian GSC, and HLM2 models are clustered together with a similar pattern close to the ratio of one. Thus, in terms of estimated SE for treatment coefficient, these three models are performing well, regardless of data type and sample size.

Treatment Coefficient: Coverage Probability. Figure 11 shows that, overall, HLM1 consistently has the lowest coverage (it falls below the lower confidence band). The rest of the models, with one exception (the Exponential GSC model for continuous Gaussian data), are all close to or above the nominal value of 95% and within the drawn confidence bands, for all data types and sample sizes.

Treatment Coefficient: Power. Figure 12 shows that the power increases as sample size increases. With a sample size of 150, all of the tests consistently have low power. Regardless of data type, HLM1 has the highest power; however, because HLM1 consistently had the lowest coverage probability it will be excluded from the present discussion. HLM2 consistently has the lowest power across all data types and sample sizes; only for sample sizes of 350 and 500 does this low power still fall within an acceptable range. Surprisingly, the Gaussian model has the highest power, regardless of data type and sample size. Although the increase in power compared to other models might be small, this effect can be more pronounced for smaller sample sizes. In general, all models have acceptable mean power for Treatment coefficient for sample sizes of 350 and 500 regardless of data type. Although sample sizes of 500 lead to extremely over-powered tests, being over-powered is not necessarily problematic if one can separate the practical importance from the statistical importance.

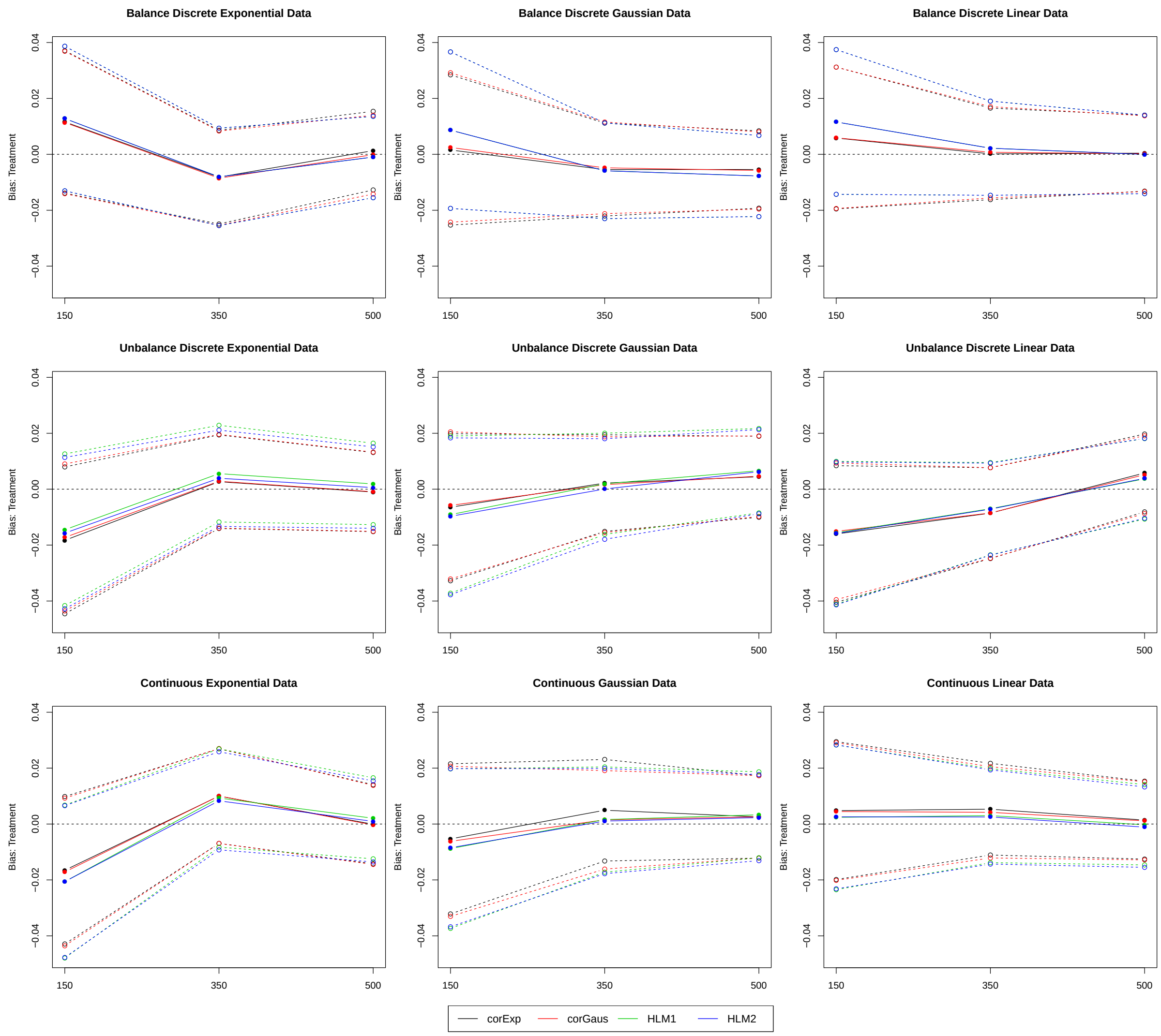


Figure 9. *Bias for the treatment coefficient with different data specifications*

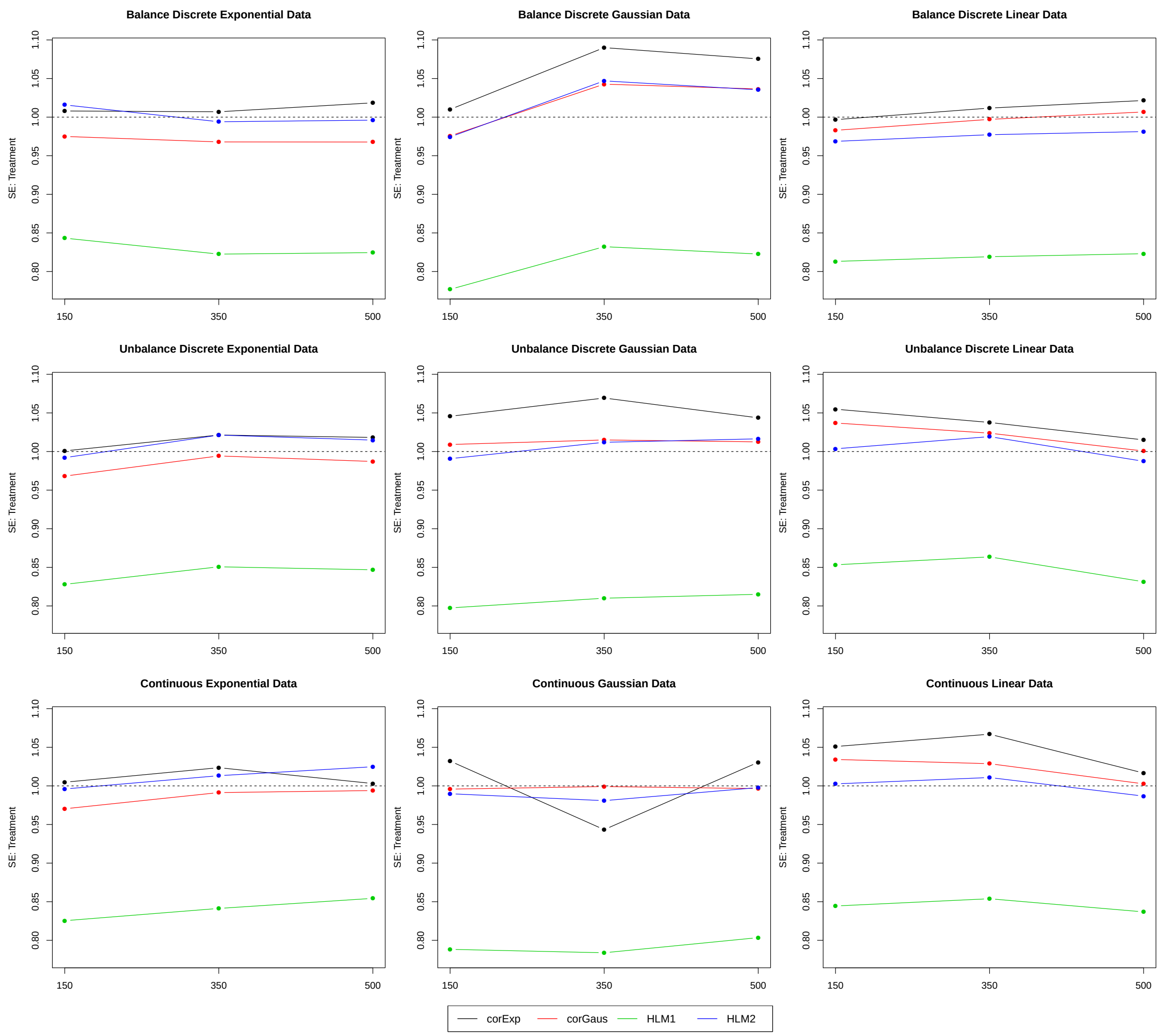


Figure 10. *Standard error for the treatment coefficient with different data specifications*

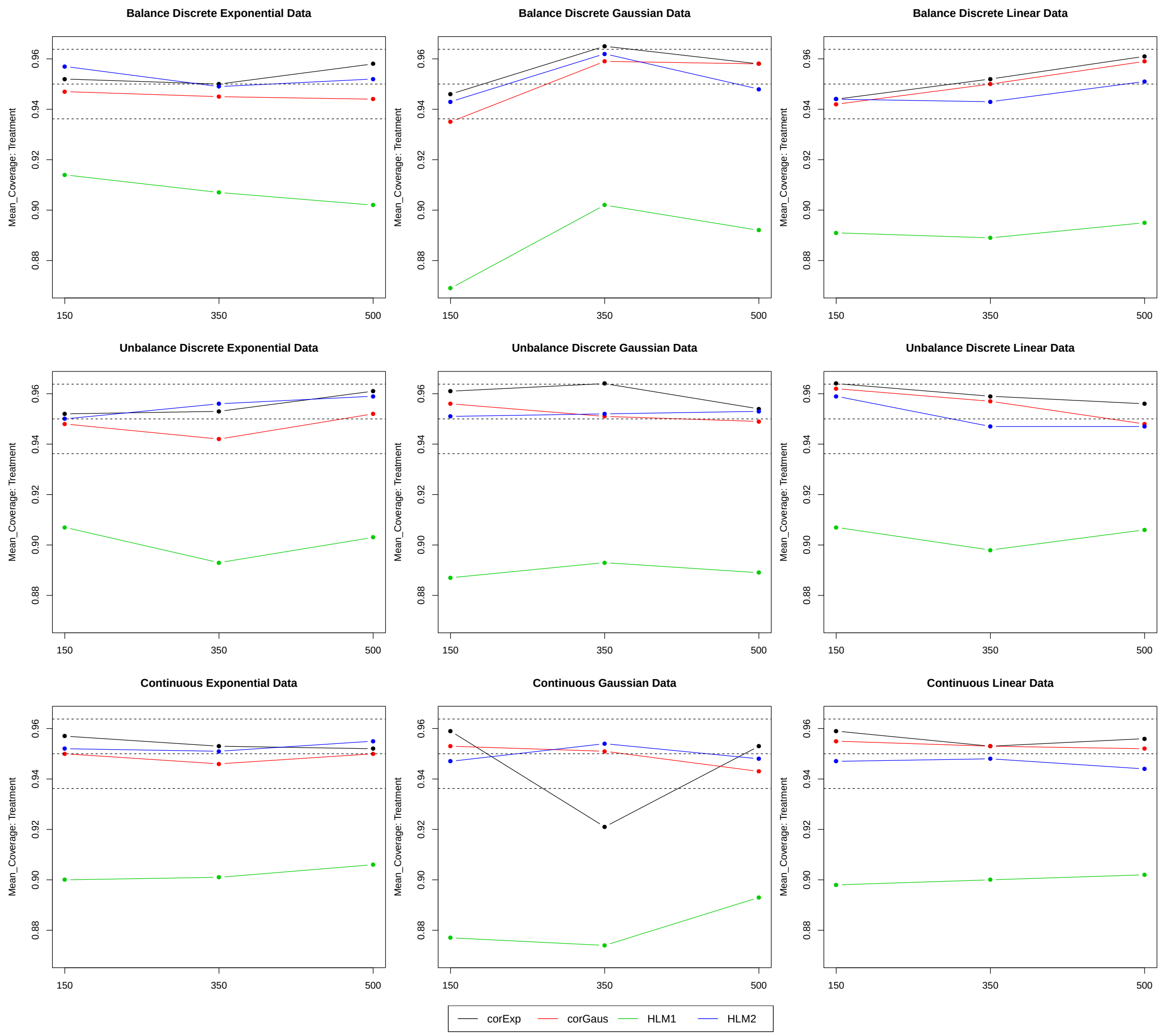


Figure 11. *Mean coverage for the treatment coefficient with different data specifications*

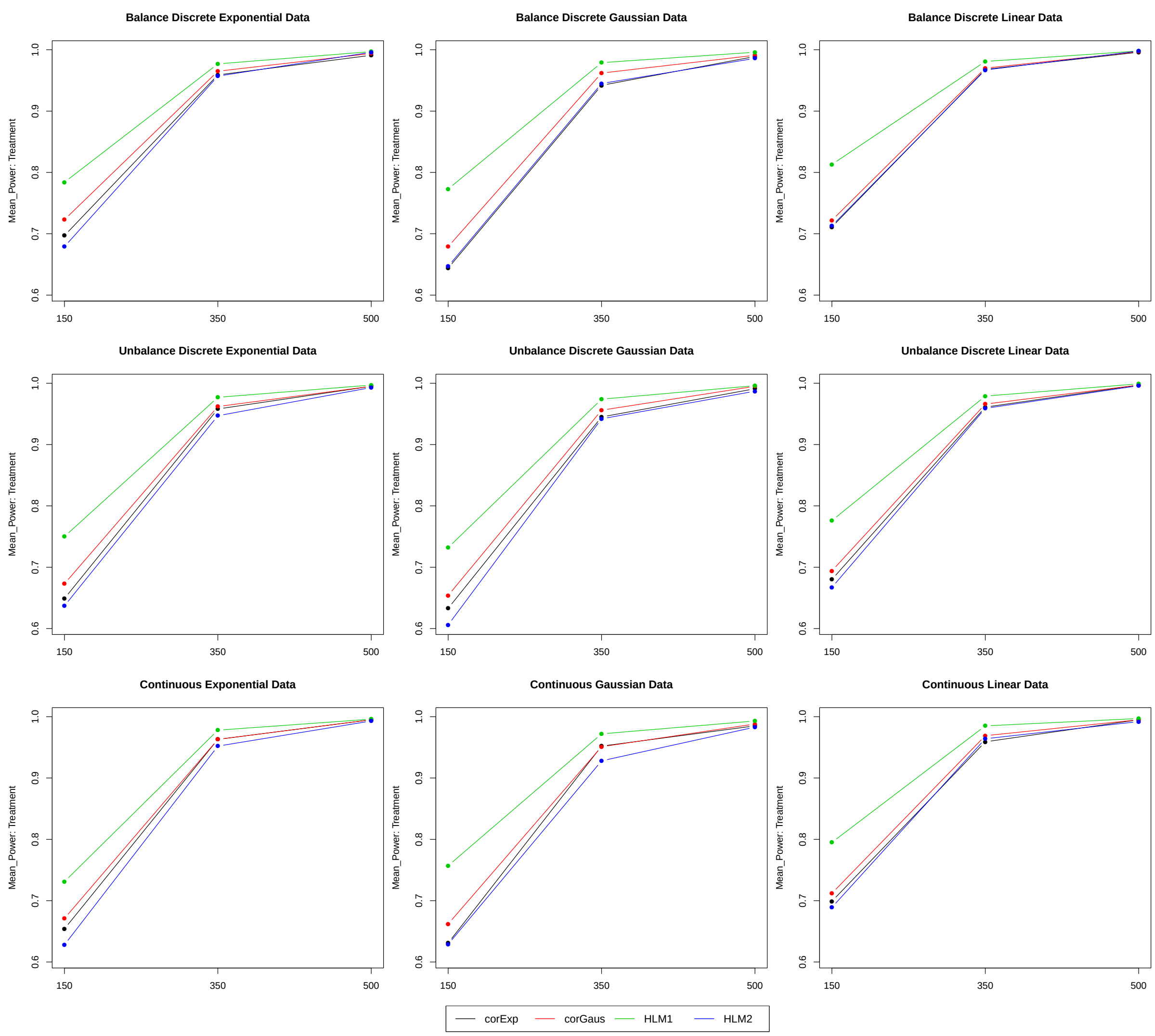


Figure 12. *Mean power for the treatment coefficient with different data specifications*

Time by Treatment Coefficient Simulation Results

Time by Treatment Coefficient: Bias. Figure 13 shows that all of the interaction coefficients are unbiased and within the shown confidence bands. Note that the observed increases/decreases are not as pronounced as they appear in the plots due to the very small range of the y-axes. Thus, all estimates of interaction coefficients are unbiased, regardless of data type and sample size.

Time by Treatment Coefficient: Standard Error. Figure 14 shows that HLM1 consistently has the smallest SE ratio (≤ 1), regardless of data type and sample size; this means that again HLM1's estimated SE is smaller than the "true" SE for the interaction coefficient. However, the other three models' SE ratio estimates are close to one, which means all the estimates are close to the "true" SE, regardless of data type and sample size. Surprisingly, HLM2's SE ratio estimates are very close to one regardless of data type and sample size. In general, the GSC models and HLM2 all have acceptable estimated SE for the interaction term.

Time by Treatment Coefficient: Coverage Probability. Figure 15 shows that HLM1 consistently has the lowest mean coverage. However, all models (except HLM1) either reach the nominal coverage probability of 95% or are very close to it, regardless of data type and sample size. Ignoring the HLM1 model, the Exponential GSC model almost consistently has the highest coverage probability, with the only exception being continuous Exponential data (which the HLM2 model is barely outperforming). In general, all models except HLM1 demonstrate adequate coverage probability for the interaction term, regardless of data type and sample size.

Time by Treatment Coefficient: Power. Figure 16 shows that for the power of the interaction term, there is an upward trend as sample size increases. With sample sizes of 350 and 500, high power is observed. Again, HLM1 will not be included in the discussion in this section since it did not have appropriate power. The Gaussian GSC model usually has the highest power and HLM2 almost consistently has the lowest. Overall, all models have acceptable power for interaction term for sample sizes of 350 and 500. Even for sample sizes of 150, the lowest power for interaction term is not severely under-powered.

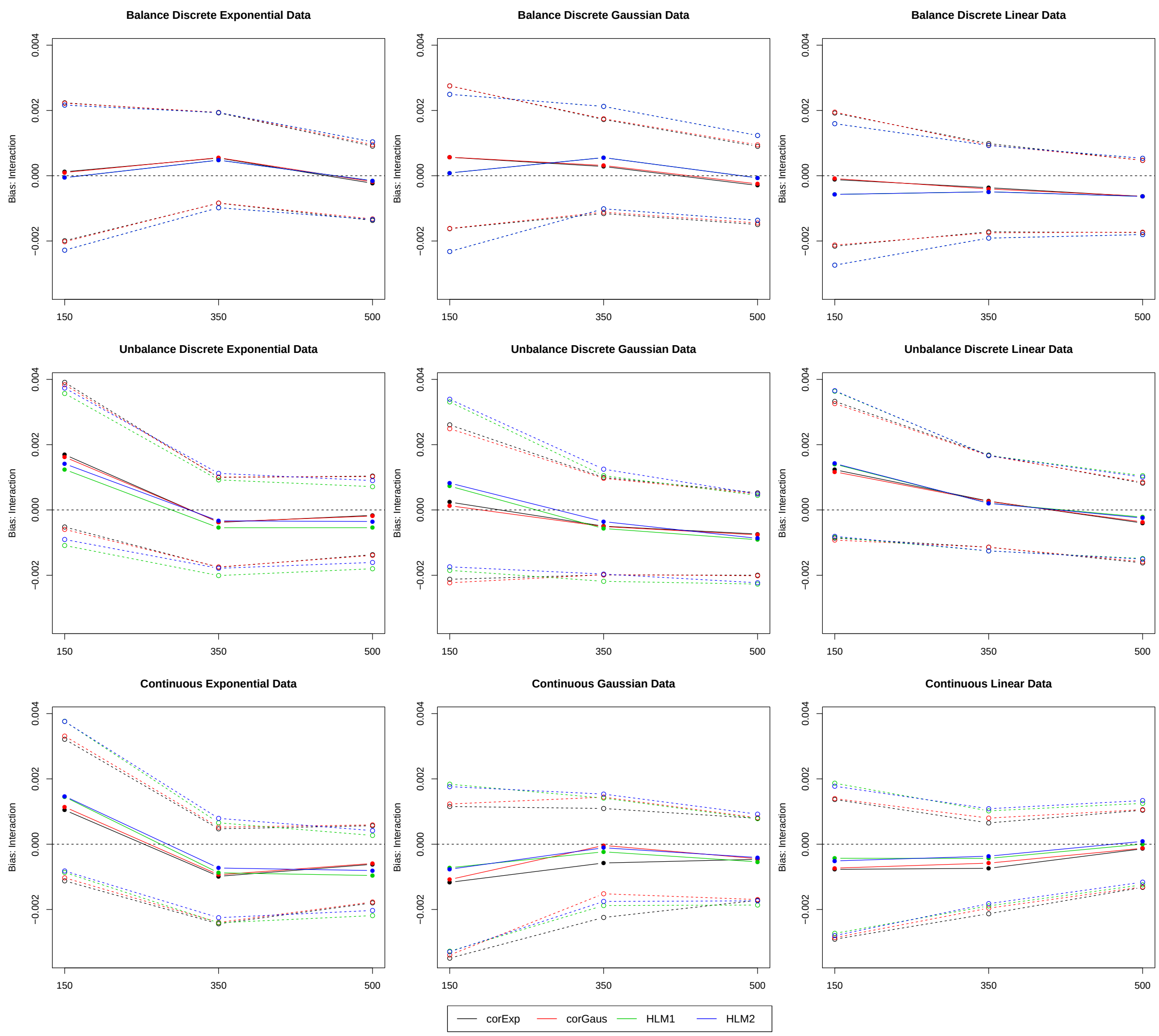


Figure 13. *Bias for the interaction between time and treatment coefficient with different data specifications*

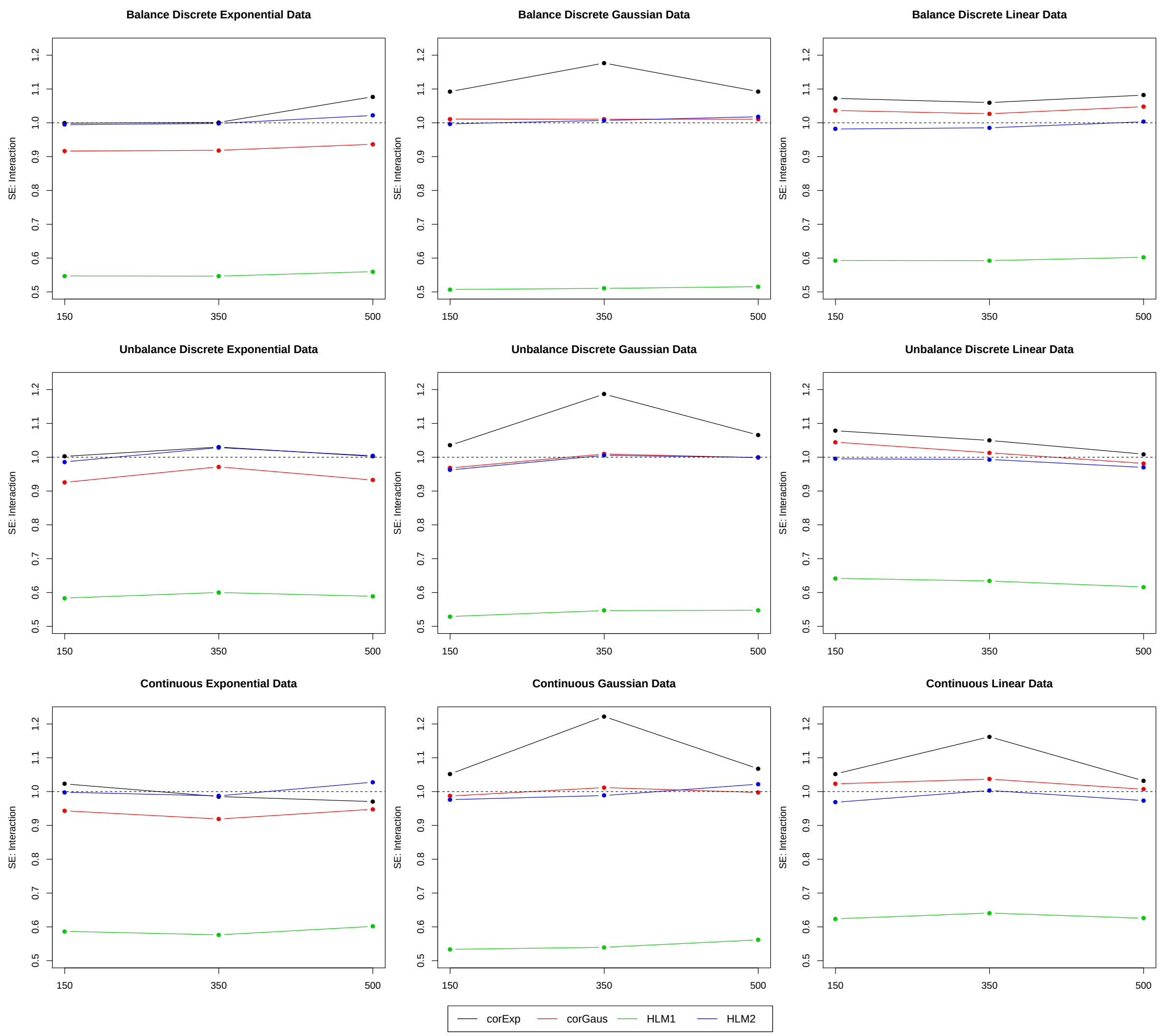


Figure 14. *Standard error for the interaction between time and treatment coefficient with different data specifications*

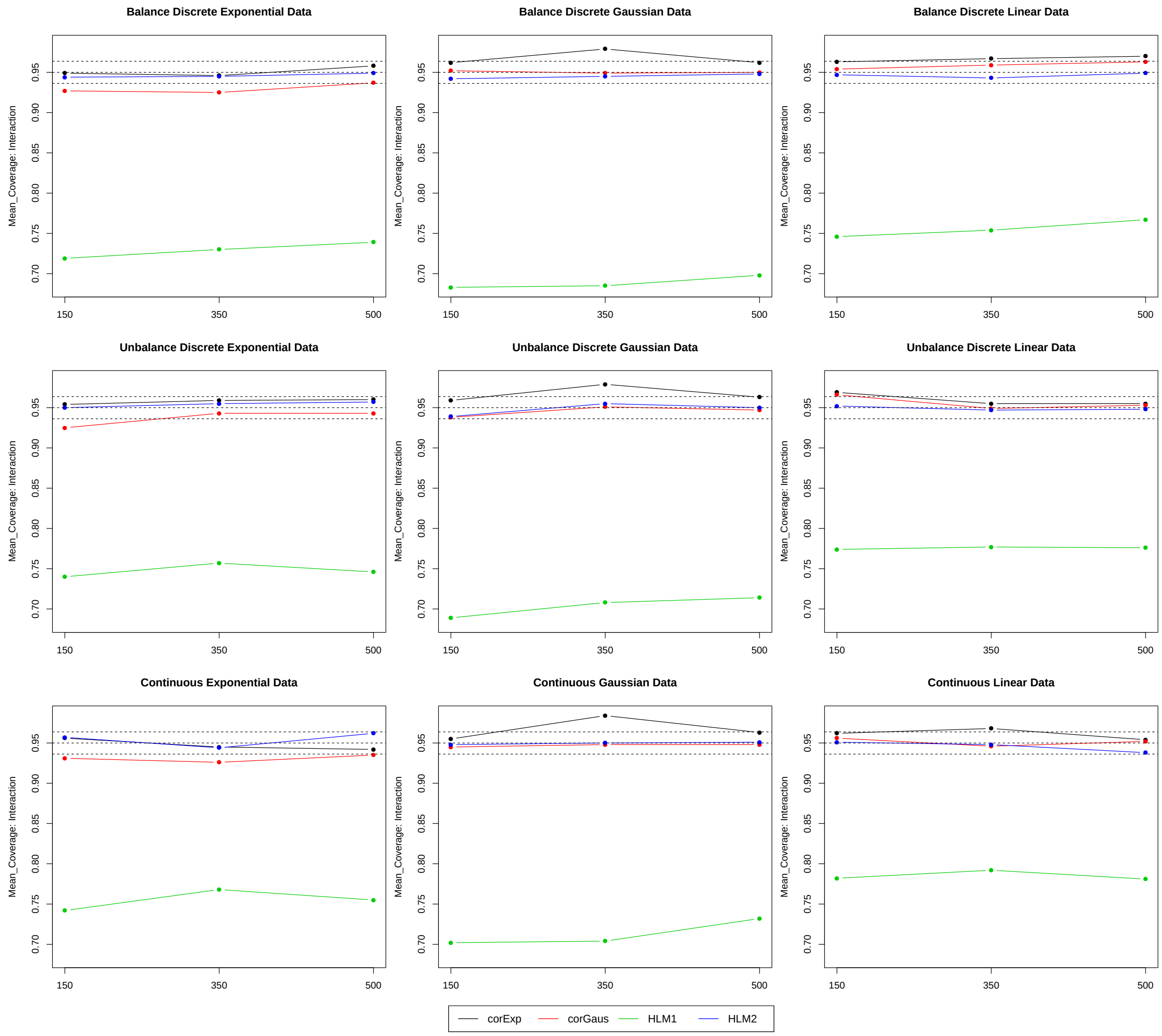


Figure 15. *Mean coverage for the interaction between time and treatment coefficient with different data specifications*

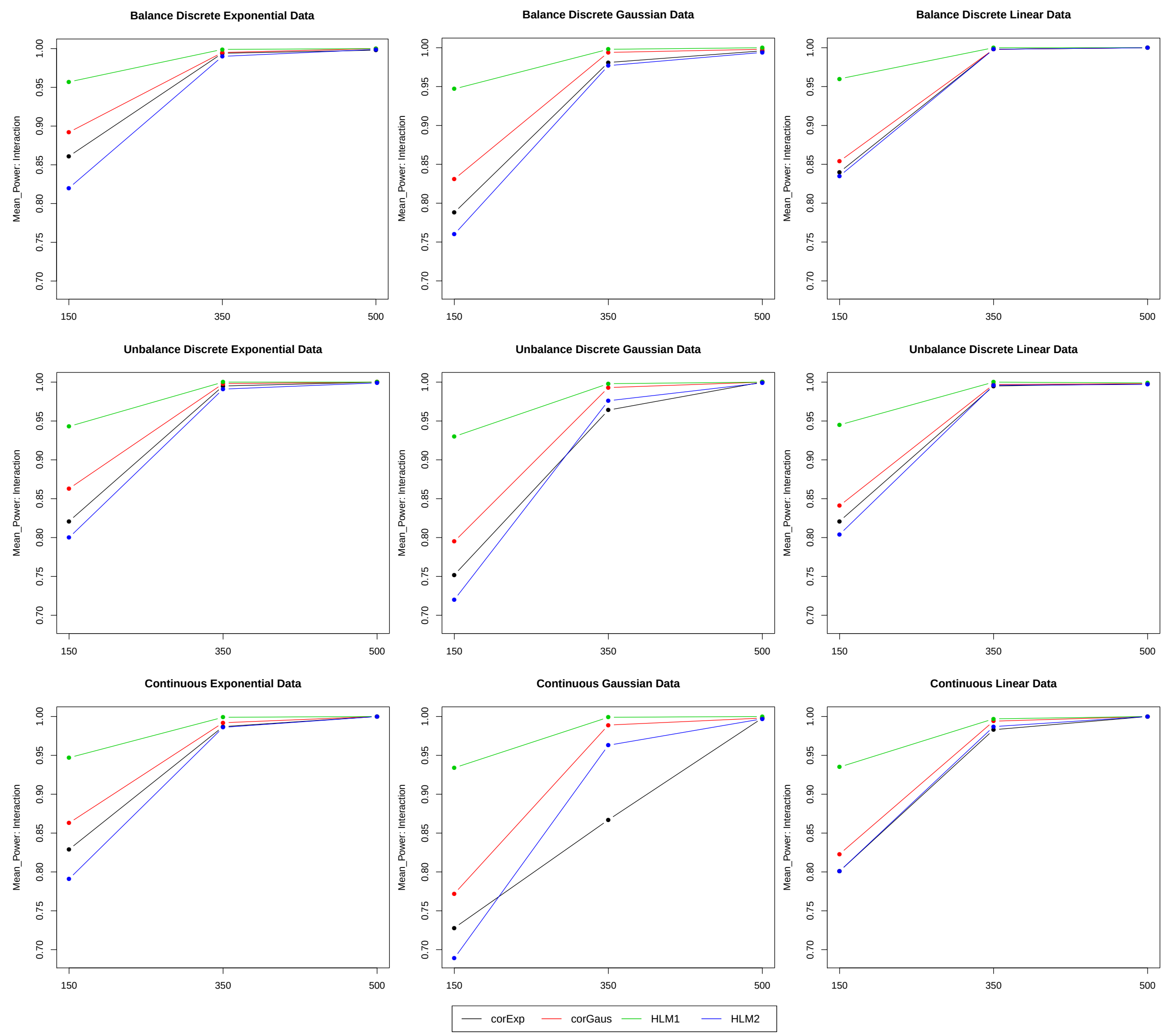


Figure 16. *Mean power for the interaction between time and treatment coefficient with different data specifications*

Model Convergence Issues in R

When running this simulation, one of the most time consuming issues was the problem of non-convergence with parameter estimation in R. The non-convergence issues were partially alleviated by using the `"L-BFGS-B"` optimization method (explained in the Simulation Specifications section) and skipping the leftover non-convergence seeds. Table 2 shows the number of problematic seeds, per 1000 simulations, categorized by model and data type.

HLM1 and HLM2 have no problematic seeds. The GSC model with an Exponential covariance structure has only 38 problematic seeds out of the total of 27000, which makes it a stable GSC model. On the other hand, the Gaussian GSC model has 512 problematic seeds, however 267 of these are produced by one data type, namely, bal_dis_n500_linear_data which is a mismatch between the data simulation process and model fitting. Ignoring this data type, the Gaussian GSC model has only 245 problematic seeds out of 27000 data sets, an acceptable non-convergence rate. Note that when this issue were explored by increasing the parameter d in the Linear GSC data, the number of problematic seeds for bal_dis_n500_linear_data decreased to double digits. This convergence issue could therefore be due to the parameter d being too small in Linear data for the Gaussian GSC model to converge. Thus, this simulation study shows that Gaussian and Exponential GSC models are, on the whole, stable spatial correlation models.

Note that according to Table 2, when the covariance structure from the data matches the models, there exist a very low number of problematic seeds per 1000 simulations.

Additionally, the effect of increasing the parameter d in data simulations of a Linear covariance structure was briefly explored. The results were not particularly sensitive to increases in the value of d (at least across the different data types), and the number of problematic seeds sporadically increased and decreased among different data types. However, for sample sizes of 500, there was a consistent decrease in the problematic number of seeds when d increased. Note that in reality, adjusting the parameter d is not an option since this value is dictated by the data. That said, it is important to observe that convergence issues can increase or decrease depending on data type.

Finally, to explore the effect of sample size on convergence issues in R, the problematic seed counts by sample size and data type were tabulated; the results are shown in Table 3. No pronounced systematic increase or decrease was observed.

Discussion

It was surprising to observe that with spatial correlation structures, running the model that matches the data simulation process (i.e. having the “right" model) did not always produce the best estimation of bias, SE, coverage probability, and power. This is an interesting and positive observation in the sense that if the data in hand is consistent with the GSC model with spatial correlation structure, using a partially mis-specified model is not a serious concern as long as the researcher chooses the right category of model. The following interesting patterns were observed comparing the four models of interest, namely, the Gaussian and Exponential GSC models, the random intercept-only model (HLM1), and the random intercept and random slope model (HLM2):

- All the estimates were unbiased, regardless of data type, sample size, and choice of modeling.
- In terms of SE, HLM1 consistently had the estimated SE the furthest below the “true” value, regardless of data type and sample size. The estimated SEs for the rest of the models were very close to the “true” SE, regardless of data type and sample size.
- In terms of coverage probability, HLM1 was outside the confidence band for all the coeffi-

	GSC Exp Model	GSC Gaus Model	HLM1 Model	HLM2 Model
bal_dis_n150_gaussian	0	0	0	0
bal_dis_n150_exponential	2	16	0	0
bal_dis_n150_linear	3	6	0	0
bal_dis_n350_gaussian	3	1	0	0
bal_dis_n350_exponential	0	12	0	0
bal_dis_n350_linear	0	11	0	0
bal_dis_n500_gaussian	0	0	0	0
bal_dis_n500_exponential	0	63	0	0
bal_dis_n500_linear	0	267	0	0
unbal_dis_n150_gaussian	0	6	0	0
unbal_dis_n150_exponential	0	8	0	0
unbal_dis_n150_linear	0	0	0	0
unbal_dis_n350_gaussian	0	0	0	0
unbal_dis_n350_exponential	18	1	0	0
unbal_dis_n350_linear	1	1	0	0
unbal_dis_n500_gaussian	0	0	0	0
unbal_dis_n500_exponential	0	11	0	0
unbal_dis_n500_linear	0	11	0	0
cts_n150_gaussian	1	14	0	0
cts_n150_exponential	0	8	0	0
cts_n150_linear	0	8	0	0
cts_n350_gaussian	5	0	0	0
cts_n350_exponential	0	21	0	0
cts_n350_linear	1	5	0	0
cts_n500_gaussian	0	0	0	0
cts_n500_exponential	4	41	0	0
cts_n500_linear	0	1	0	0
Total Problematic Seeds	38	512	0	0

Table 2. *Number of problematic seeds per data set by model*

Note: Here bal="balanced", unbal="unbalanced", cts="continuous", dis="discrete", n150="sample size of 150" (similar notation is used for sample sizes of 350 and 500). Linear, exponential and gaussian denote the implemented covariance structure of the data. GSC Exp and GSC Gaussian represent the GSC model with Exponential and Gaussian covariance structure, respectively. HLM1 denotes the random intercept only model. HLM2 denotes the random intercept and slope model.

	Balance Discrete			**Unbalance Discrete**			**Unbalance Continuous**		
Sample Size	150	350	500	150	350	500	150	350	500
Exponential GSC Data	18	12	63	8	19	11	8	21	45
Gaussian GSC Data	0	4	0	6	0	0	15	5	0
Linear GSC Data	9	11	267	0	2	11	8	6	1

Table 3. *Number of problematic seeds per data set by model, separated by sample size*

cients, regardless of data type and sample size. The rest of the models were very close or within the confidence band for coverage probability.

- Of those with adequate coverage, the GSC Gaussian model had slightly higher power than the others. HLM2 almost consistently had the lowest power (although still in acceptable range) for all the coefficients, regardless of data type and sample size.

According to the simulation study, researchers can best estimate bias, SE, coverage probability, and power by choosing a sample size of 350 to 500 for individual repeated measures of 10 to 15 (assuming resource allocation and ethical considerations are not issues). However, for a given sample size of 150 or lower, researchers must be more cautious when selecting the covariance structure to assure reliable inference and estimation.

In terms of choice of modeling, HLM1 consistently has the lowest coverage probability, making it one of the low performing models. On the other hand, the HLM2 and the GSC Gaussian/Exponential models perform well for all of the coefficients with sample sizes of 350 and 500, regardless of data type. The GSC model with a Gaussian covariance structure has the highest power for all coefficients, regardless of data type and sample size. This high performance is especially notable for extending to the smallest sample size of 150. Conversely, HLM2 consistently has the lowest power. Although HLM2's relatively low power was generally in acceptable range, this effect can be magnified with sample sizes smaller than 150. Thus, because it is likely that researchers are working with a sample size of 150 and possibly smaller, choosing either Gaussian GSC, Exponential GSC, or HLM2 would make sense in most scenarios, keeping in mind that HLM2 has the lowest power among these models and a possibly unrealistic covariance pattern, derived in Appendix C.

In choosing the best model, researchers working with a preset sample size should keep in mind the assumptions that all these models make by implementing different covariance structures.

For example, HLM1 induces CS, which is the only covariance model that is both an LMM and a covariance pattern model. Thus, in HLM1, as in CS, distance in time (i.e time lag) is not taken into account and the model assumes equal measurement intervals with no missingness; with real data, these assumptions are usually not met. Similarly, AR(1) needs equal time lags (note: in the simulated balanced discrete data set, AR(1) is the same as Exponential covariance structure). Due to these unrealistic assumptions and the inferior performance of HLM1 in terms of coverage probability, researchers should be cautious when choosing a random intercept model.

HLM2, the GSC Gaussian, and Exponential models can all handle unequally spaced data and distinguish within- and between-variability. HLM2 did consistently have the lowest power, though HLM2's power generally remains in an acceptable range even for this simulation study's smallest sample size of 150. The problem of low power might be more pronounced with smaller sample sizes.

Moreover, in HLM2, the implementation of random effects (i.e. random intercept and random slope) induces a more difficult correlation structure than spatial correlation structures. Mathematically speaking, Appendix C shows the derivation of the variance/covariance structures of the GSC and HLM2 models. Looking at equations 6 and 7, which correspond to the variance and covariance of the GSC model, one can observe that if the Gaussian serial correlation is chosen, these equations result in a decrease in correlation as time lag increases. However, equations 8 and 9 show that the variance and covariance of the HLM2 model both most of the time increases as time lag increases. In general, this assumption is not realistic and needs to be closely examined before choosing HLM2.

Overall, judging based on bias, SE, coverage probability, and power, HLM1 was the lowest performing model, though the three other models presented in this simulation study performed well. The performances of HLM2, Exponential GSC, and Gaussian GSC models were all acceptable, with HLM2 having the lowest power (although still in an acceptable range) and the Gaussian GSC having the highest power (since the HLM1 was already discarded as one of the lowest performing models).

Conclusion

Commonly used multilevel modeling approaches in longitudinal research such as HLM/LMM can be improved in Education and Psychology by taking the additional step of modeling the covariance structure. The GSC model can be defined as an extension of LMM where one inserts the covariance pattern into the modeling process. As shown in the first paper, this model is very under-used in Education and Psychology and would be a useful addition to the longitudinal literature in these fields.

The focus of this paper was to introduce the GSC model to the fields of Education and Psychology. A simulation study was performed to investigate how the GSC model might improve on regularly used basic HLM methods when the data are consistent with the GSC model with spatial covariance structures. The simulation study itself had two main purposes. First, the *estimation* properties of the fixed effect were explored by looking at bias and standard error of estimates. Second, the *testing* properties were examined by looking into the coverage probability of 95% confidence interval and the power of the Wald test to detect meaningful difference. The under-used GSC models with spatial covariance patterns (i.e. Exponential and Gaussian) were compared to standard HLM models with random intercept (HLM1) and random intercept/slope (HLM2) using data consistent with the GSC model with Exponential, Gaussian, and Linear covariance structure.

The simulation results can be summarized in four main points. First, in terms of bias, as expected, all models produced unbiased estimates. Second, unless there is strong evidence that HLM1 should be used, it should be disregarded as a modeling choice (specially when spatial covariance structure is observed) since it had the lowest coverage probability. Third, the GSC model with Gaussian and Exponential covariance structures were the most stable models, with good testing and estimation properties regardless of sample size and data type. Fourth, HLM2 can be chosen as the third best model, however it consistently had the lowest power (although still in an acceptable range) and this effect can be magnified with fewer repeated measures and smaller sample sizes. Thus, researchers in Education and Psychology can greatly benefit from employing, when appropriate, the GSC model with Gaussian and Exponential covariance patterns. Simple exploratory tools such as the variogram can assist researchers in determining when these GSC models are suitable. To introduce this tool to researchers in Education and Psychology, the final paper of this dissertation will include a tutorial on how to use variograms to select the best model.

Despite its unrealistic covariance structure (derived in Appendix C), the relatively strong performance of the HLM2 model compared to the GSC Exponential and Gaussian methods merits further exploration. HLM2 consistently had the smallest power, even though it was unbiased and its SE was generally close to the “true” SE; this observation should be investigated further in future drafts of this dissertation.

In this simulation study, the effect of choosing the “right" model was not explored in terms of the estimation of covariance structure. Although all of the estimations of the covariance structures have been saved, the task of testing and exploring the covariance parameters is reserved for another paper.

Exploring R's convergence issue for the "right" and "wrong" models is another subject that deserves its own study. Although R convergence issues have been explored to some extent, the mathematics of the likelihood functions and the connection with R's optimization process are two areas only briefly discussed here; these subjects merit further attention.

All of the estimations in this study were based on Maximum Likelihood (ML) estimates; the use of Restricted Maximum Likelihood (REML) was not explored. Comparing the estimation and testing properties of ML versus REML for different models and sample sizes would be beneficial for future research. Additional avenues of study related to this paper might further investigate (a) how smaller effect sizes affect estimating power and (b) the feasibility of smaller sample sizes in combination with differing numbers of repeated measures.

Finally, instead of exploring *all* of the plotted simulation results visually, one could use ANOVA and logistic regression to identify the statistical significance of specific plots using interaction terms. These methods entail exploring the statistical significance of the four-way interaction between covariance structure of data, sample size, choice of model, and data type with all the lower level interactions. Note that these methods would use bias, SE, coverage probability, or power as an outcome measure. Due to time constraints, this analysis was not performed, but this option was considered in detail and remains an interesting topic that deserves more attention.

References

Barnett, A. G., Koper, N., Dobson, A. J., Schmiegelow, F., & Manseau, M. (2010). Using information criteria to select the correct variance–covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution*, *1*(1), 15–24.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208.

DeGraff, J. V., DeGraff, N., & Romesburg, H. C. (2013). Literature searches with Google Scholar: Knowing what you are and are not getting. *GSA Today*, *23*(10), 44–45.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics*, 959–971.

Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we preach: Turning tables into graphs. *The American Statistician*, *56*(2), 121–130.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics-Simulation and computation*, *27*(3), 591–604.

Kwok, O.-m., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research*, *42*(3), 557–592.

Martin-Martin, A., Orduna-Malea, E., Harzing, A.-W., & López-Cózar, E. D. (2017). Can we use Google Scholar to identify highly-cited documents? *Journal of Informetrics*, *11*(1), 152–163.

Martinussen, T., Skovgaard, I. M., & Sorensen, H. (2012). *A first guide to statistical computations in r*. Samfundslitteratur.

Nocedal, J., & Wright, S. J. (1999). *Springer series in operations research. numerical optimization.* New York, NY: Springer.

Pusponegoro, N. H., Notodiputro, K. A., Sartono, B., et al. (2017). Linear mixed model for analyzing longitudinal data: A simulation study of children growth differences. *Procedia Computer Science*, *116*, 284–291.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org`

Rochon, J. (1991). Sample size calculations for two-group repeated-measures experiments. *Biometrics*, 1383–1398.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Willett, J. B. (1988). Chapter 9: Questions and answers in the measurement of change. *Review of Research in Education*, *15*(1), 345–422.

Appendices

Appendix A

Exponential to AR(1) Equivalency Derivation

The AR(1) covariance structure is here defined as follows:

$$\rho(u) = \psi^u \tag{3}$$

where $u = |t_{ij} - t_{ik}|$ which is the time lag between the two different measurements for the same subject. Note that usually the Greek letter ρ is used for AR(1) parameter, but to prevent confusion with the name of the serial correlation function ψ was used for the AR(1) correlation parameter. The Exponential covariance structure is here defined as follows:

$$\rho(u) = e^{-\frac{u}{\phi}} \tag{4}$$

Then equation 3 is equivalent to equation 4 such that:

$$\rho(u) = e^{-\frac{u}{\phi}} = (e^{-\frac{1}{\phi}})^u = \psi^u \iff \psi = e^{-\frac{1}{\phi}} \tag{5}$$

Appendix B
Numerical Results

Note that for all the tables in this appendix (i.e. Tables B1 to B27) the column names are defined as follow:

- corExp denotes the GSC model with Exponential covariance structure.
- corGaus denotes the GSC model with Gaussian covariance structure.
- HLM1 denotes the random intercept only model.
- HLM2 denotes the random intercept and slope model.

Each row name represents the name of the parameter followed by the name of estimation. For example, "Time: Bias" means the estimation of the bias for the Time parameter. The respective data type is written in each table caption.

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0009	-0.0021	-0.0085	-0.0085
Intercept: SE of Coefficient	0.3034	0.2916	0.2431	0.3049
Intercept: Coverage Probability	0.9530	0.9410	0.8680	0.9500
Intercept: Power	0.9050	0.9150	0.9440	0.8880
Time: Bias	-0.0007	-0.0006	-0.0000	-0.0000
Time: SE of Coefficient	0.0267	0.0247	0.0137	0.0268
Time: Coverage Probability	0.9630	0.9440	0.6730	0.9480
Time: Power	0.9720	0.9840	0.9970	0.9680
Treatment: Bias	0.0016	0.0025	0.0087	0.0087
Treatment: SE of Coefficient	0.4291	0.4123	0.3439	0.4312
Treatment: Coverage Probability	0.9460	0.9350	0.8690	0.9430
Treatment: Power	0.6440	0.6790	0.7730	0.6470
Time by Treatment Interaction: Bias	0.0006	0.0006	0.0001	0.0001
Time by Treatment Interaction: SE of Coefficient	0.0378	0.0349	0.0193	0.0380
Time by Treatment Interaction: Coverage Probability	0.9620	0.9520	0.6830	0.9420
Time by Treatment Interaction: Power	0.7880	0.8310	0.9470	0.7600

Table B1. *Balanced discrete data with sample size 150 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0031	-0.0033	-0.0051	-0.0051
Intercept: SE of Coefficient	0.2870	0.2777	0.2438	0.2937
Intercept: Coverage Probability	0.9500	0.9410	0.8850	0.9460
Intercept: Power	0.9310	0.9380	0.9530	0.9200
Time: Bias	-0.0003	-0.0002	-0.0000	-0.0000
Time: SE of Coefficient	0.0236	0.0217	0.0136	0.0247
Time: Coverage Probability	0.9510	0.9240	0.7150	0.9430
Time: Power	0.9890	0.9910	0.9980	0.9860
Treatment: Bias	0.0116	0.0114	0.0128	0.0128
Treatment: SE of Coefficient	0.4059	0.3927	0.3448	0.4154
Treatment: Coverage Probability	0.9520	0.9470	0.9140	0.9570
Treatment: Power	0.6970	0.7230	0.7840	0.6790
Time by Treatment Interaction: Bias	0.0001	0.0001	-0.0001	-0.0001
Time by Treatment Interaction: SE of Coefficient	0.0334	0.0307	0.0192	0.0349
Time by Treatment Interaction: Coverage Probability	0.9490	0.9270	0.7190	0.9440
Time by Treatment Interaction: Power	0.8610	0.8920	0.9570	0.8200

Table B2. *Balanced discrete data with sample size 150 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0018	-0.0018	-0.0056	-0.0056
Intercept: SE of Coefficient	0.2826	0.2778	0.2353	0.2802
Intercept: Coverage Probability	0.9380	0.9330	0.8770	0.9250
Intercept: Power	0.9320	0.9410	0.9560	0.9300
Time: Bias	-0.0005	-0.0006	-0.0002	-0.0002
Time: SE of Coefficient	0.0244	0.0236	0.0144	0.0238
Time: Coverage Probability	0.9520	0.9460	0.7270	0.9310
Time: Power	0.9810	0.9840	0.9980	0.9770
Treatment: Bias	0.0058	0.0059	0.0116	0.0116
Treatment: SE of Coefficient	0.3997	0.3929	0.3327	0.3963
Treatment: Coverage Probability	0.9440	0.9420	0.8910	0.9440
Treatment: Power	0.7110	0.7220	0.8130	0.7130
Time by Treatment Interaction: Bias	-0.0001	-0.0001	-0.0006	-0.0006
Time by Treatment Interaction: SE of Coefficient	0.0346	0.0334	0.0203	0.0336
Time by Treatment Interaction: Coverage Probability	0.9630	0.9540	0.7460	0.9470
Time by Treatment Interaction: Power	0.8400	0.8540	0.9600	0.8350

Table B3. *Balanced discrete data with sample size 150 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0076	0.0062	0.0037	0.0037
Intercept: SE of Coefficient	0.2025	0.1911	0.1593	0.2004
Intercept: Coverage Probability	0.9740	0.9670	0.9030	0.9650
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0007	-0.0007	-0.0005	-0.0005
Time: SE of Coefficient	0.0190	0.0162	0.0090	0.0177
Time: Coverage Probability	0.9760	0.9460	0.6920	0.9480
Time: Power	0.9990	0.9990	1.0000	0.9980
Treatment: Bias	-0.0055	-0.0048	-0.0059	-0.0059
Treatment: SE of Coefficient	0.2863	0.2702	0.2253	0.2835
Treatment: Coverage Probability	0.9650	0.9590	0.9020	0.9620
Treatment: Power	0.9420	0.9620	0.9790	0.9450
Time by Treatment Interaction: Bias	0.0003	0.0003	0.0006	0.0006
Time by Treatment Interaction: SE of Coefficient	0.0268	0.0229	0.0127	0.0250
Time by Treatment Interaction: Coverage Probability	0.9790	0.9490	0.6850	0.9450
Time by Treatment Interaction: Power	0.9810	0.9940	0.9980	0.9770

Table B4. *Balanced discrete data with sample size 350 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0088	0.0090	0.0073	0.0073
Intercept: SE of Coefficient	0.1885	0.1824	0.1602	0.1936
Intercept: Coverage Probability	0.9450	0.9400	0.8990	0.9480
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0007	-0.0007	-0.0005	-0.0005
Time: SE of Coefficient	0.0155	0.0143	0.0089	0.0163
Time: Coverage Probability	0.9470	0.9210	0.7220	0.9450
Time: Power	1.0000	1.0000	1.0000	0.9990
Treatment: Bias	-0.0082	-0.0085	-0.0081	-0.0081
Treatment: SE of Coefficient	0.2666	0.2579	0.2266	0.2738
Treatment: Coverage Probability	0.9500	0.9450	0.9070	0.9490
Treatment: Power	0.9590	0.9650	0.9770	0.9570
Time by Treatment Interaction: Bias	0.0005	0.0006	0.0005	0.0005
Time by Treatment Interaction: SE of Coefficient	0.0219	0.0202	0.0126	0.0230
Time by Treatment Interaction: Coverage Probability	0.9460	0.9250	0.7300	0.9450
Time by Treatment Interaction: Power	0.9940	0.9950	0.9990	0.9900

Table B5. *Balanced discrete data with sample size 350 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0020	0.0012	-0.0014	-0.0014
Intercept: SE of Coefficient	0.1852	0.1820	0.1542	0.1840
Intercept: Coverage Probability	0.9580	0.9590	0.8940	0.9550
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0002	-0.0002	0.0001	0.0001
Time: SE of Coefficient	0.0160	0.0154	0.0094	0.0156
Time: Coverage Probability	0.9670	0.9630	0.7610	0.9510
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0002	0.0007	0.0022	0.0022
Treatment: SE of Coefficient	0.2619	0.2575	0.2181	0.2602
Treatment: Coverage Probability	0.9520	0.9500	0.8890	0.9430
Treatment: Power	0.9680	0.9700	0.9810	0.9670
Time by Treatment Interaction: Bias	-0.0004	-0.0004	-0.0005	-0.0005
Time by Treatment Interaction: SE of Coefficient	0.0226	0.0218	0.0133	0.0221
Time by Treatment Interaction: Coverage Probability	0.9670	0.9590	0.7540	0.9430
Time by Treatment Interaction: Power	0.9980	0.9980	1.0000	0.9980

Table B6. *Balanced discrete data with sample size 350 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0082	0.0081	0.0075	0.0075
Intercept: SE of Coefficient	0.1663	0.1600	0.1335	0.1680
Intercept: Coverage Probability	0.9630	0.9590	0.8890	0.9560
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0001	-0.0001	-0.0000	-0.0000
Time: SE of Coefficient	0.0146	0.0135	0.0075	0.0148
Time: Coverage Probability	0.9630	0.9520	0.6870	0.9550
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	-0.0054	-0.0058	-0.0078	-0.0078
Treatment: SE of Coefficient	0.2352	0.2263	0.1888	0.2376
Treatment: Coverage Probability	0.9580	0.9580	0.8920	0.9480
Treatment: Power	0.9890	0.9910	0.9960	0.9860
Time by Treatment Interaction: Bias	-0.0003	-0.0003	-0.0001	-0.0001
Time by Treatment Interaction: SE of Coefficient	0.0207	0.0191	0.0106	0.0209
Time by Treatment Interaction: Coverage Probability	0.9620	0.9500	0.6980	0.9480
Time by Treatment Interaction: Power	0.9960	0.9980	1.0000	0.9940

Table B7. *Balanced discrete data with sample size 500 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0028	0.0036	0.0033	0.0033
Intercept: SE of Coefficient	0.1601	0.1526	0.1341	0.1620
Intercept: Coverage Probability	0.9600	0.9500	0.9010	0.9570
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0002	-0.0002	-0.0000	-0.0000
Time: SE of Coefficient	0.0137	0.0119	0.0074	0.0136
Time: Coverage Probability	0.9570	0.9270	0.7230	0.9580
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0013	-0.0001	-0.0010	-0.0010
Treatment: SE of Coefficient	0.2265	0.2158	0.1897	0.2292
Treatment: Coverage Probability	0.9580	0.9440	0.9020	0.9520
Treatment: Power	0.9910	0.9940	0.9970	0.9960
Time by Treatment Interaction: Bias	-0.0002	-0.0002	-0.0002	-0.0002
Time by Treatment Interaction: SE of Coefficient	0.0194	0.0168	0.0105	0.0192
Time by Treatment Interaction: Coverage Probability	0.9580	0.9370	0.7390	0.9490
Time by Treatment Interaction: Power	0.9980	1.0000	1.0000	0.9990

Table B8. *Balanced discrete data with sample size 500 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0028	0.0028	0.0025	0.0025
Intercept: SE of Coefficient	0.1551	0.1524	0.1292	0.1540
Intercept: Coverage Probability	0.9610	0.9540	0.9000	0.9550
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	0.0001	0.0002	0.0002	0.0002
Time: SE of Coefficient	0.0134	0.0129	0.0079	0.0131
Time: Coverage Probability	0.9700	0.9660	0.7830	0.9550
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0004	0.0002	-0.0000	-0.0000
Treatment: SE of Coefficient	0.2194	0.2156	0.1827	0.2178
Treatment: Coverage Probability	0.9610	0.9590	0.8950	0.9510
Treatment: Power	0.9960	0.9970	0.9980	0.9980
Time by Treatment Interaction: Bias	-0.0006	-0.0006	-0.0006	-0.0006
Time by Treatment Interaction: SE of Coefficient	0.0189	0.0182	0.0111	0.0185
Time by Treatment Interaction: Coverage Probability	0.9700	0.9630	0.7670	0.9490
Time by Treatment Interaction: Power	1.0000	1.0000	1.0000	1.0000

Table B9. *Balanced discrete data with sample size 500 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0048	0.0048	0.0064	0.0057
Intercept: SE of Coefficient	0.3064	0.2957	0.2494	0.3100
Intercept: Coverage Probability	0.9540	0.9470	0.8770	0.9480
Intercept: Power	0.9150	0.9300	0.9490	0.9010
Time: Bias	-0.0010	-0.0010	-0.0012	-0.0011
Time: SE of Coefficient	0.0272	0.0253	0.0150	0.0275
Time: Coverage Probability	0.9540	0.9420	0.7330	0.9490
Time: Power	0.9590	0.9700	0.9950	0.9500
Treatment: Bias	-0.0065	-0.0058	-0.0091	-0.0097
Treatment: SE of Coefficient	0.4348	0.4198	0.3544	0.4395
Treatment: Coverage Probability	0.9610	0.9560	0.8870	0.9510
Treatment: Power	0.6330	0.6540	0.7320	0.6060
Time by Treatment Interaction: Bias	0.0002	0.0001	0.0007	0.0008
Time by Treatment Interaction: SE of Coefficient	0.0387	0.0361	0.0216	0.0391
Time by Treatment Interaction: Coverage Probability	0.9590	0.9380	0.6890	0.9390
Time by Treatment Interaction: Power	0.7520	0.7950	0.9300	0.7200

Table B10. *Unbalanced discrete data with sample size 150 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0063	0.0065	0.0061	0.0055
Intercept: SE of Coefficient	0.2927	0.2834	0.2499	0.2993
Intercept: Coverage Probability	0.9560	0.9470	0.8970	0.9570
Intercept: Power	0.9290	0.9430	0.9600	0.9170
Time: Bias	-0.0012	-0.0012	-0.0010	-0.0010
Time: SE of Coefficient	0.0246	0.0227	0.0149	0.0254
Time: Coverage Probability	0.9530	0.9360	0.7550	0.9450
Time: Power	0.9780	0.9840	0.9970	0.9670
Treatment: Bias	-0.0183	-0.0173	-0.0145	-0.0157
Treatment: SE of Coefficient	0.4156	0.4024	0.3551	0.4246
Treatment: Coverage Probability	0.9520	0.9480	0.9070	0.9500
Treatment: Power	0.6490	0.6730	0.7500	0.6370
Time by Treatment Interaction: Bias	0.0017	0.0016	0.0012	0.0014
Time by Treatment Interaction: SE of Coefficient	0.0351	0.0324	0.0215	0.0362
Time by Treatment Interaction: Coverage Probability	0.9540	0.9250	0.7400	0.9500
Time by Treatment Interaction: Power	0.8210	0.8630	0.9430	0.8000

Table B11. *Unbalanced discrete data with sample size 150 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0116	0.0111	0.0120	0.0112
Intercept: SE of Coefficient	0.2869	0.2820	0.2423	0.2854
Intercept: Coverage Probability	0.9500	0.9470	0.8940	0.9450
Intercept: Power	0.9400	0.9450	0.9700	0.9340
Time: Bias	-0.0017	-0.0017	-0.0019	-0.0018
Time: SE of Coefficient	0.0250	0.0242	0.0158	0.0246
Time: Coverage Probability	0.9650	0.9610	0.8090	0.9570
Time: Power	0.9860	0.9870	0.9980	0.9840
Treatment: Bias	-0.0160	-0.0151	-0.0156	-0.0158
Treatment: SE of Coefficient	0.4075	0.4006	0.3446	0.4052
Treatment: Coverage Probability	0.9640	0.9620	0.9070	0.9590
Treatment: Power	0.6800	0.6940	0.7760	0.6670
Time by Treatment Interaction: Bias	0.0012	0.0012	0.0014	0.0014
Time by Treatment Interaction: SE of Coefficient	0.0357	0.0345	0.0227	0.0351
Time by Treatment Interaction: Coverage Probability	0.9690	0.9660	0.7740	0.9520
Time by Treatment Interaction: Power	0.8210	0.8410	0.9450	0.8040

Table B12. *Unbalanced discrete data with sample size 150 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0015	-0.0008	-0.0005	0.0007
Intercept: SE of Coefficient	0.2066	0.1943	0.1640	0.2036
Intercept: Coverage Probability	0.9580	0.9470	0.8850	0.9480
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0001	-0.0001	-0.0001	-0.0002
Time: SE of Coefficient	0.0198	0.0167	0.0099	0.0181
Time: Coverage Probability	0.9760	0.9500	0.7090	0.9500
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0022	0.0016	0.0019	0.0001
Treatment: SE of Coefficient	0.2922	0.2747	0.2318	0.2878
Treatment: Coverage Probability	0.9640	0.9510	0.8930	0.9520
Treatment: Power	0.9450	0.9560	0.9740	0.9420
Time by Treatment Interaction: Bias	-0.0005	-0.0005	-0.0006	-0.0004
Time by Treatment Interaction: SE of Coefficient	0.0280	0.0236	0.0140	0.0255
Time by Treatment Interaction: Coverage Probability	0.9790	0.9510	0.7080	0.9550
Time by Treatment Interaction: Power	0.9640	0.9930	0.9980	0.9760

Table B13. *Unbalanced discrete data with sample size 350 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0038	-0.0038	-0.0056	-0.0047
Intercept: SE of Coefficient	0.1911	0.1865	0.1645	0.1966
Intercept: Coverage Probability	0.9500	0.9470	0.9000	0.9570
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0003	-0.0003	-0.0002	-0.0003
Time: SE of Coefficient	0.0159	0.0150	0.0098	0.0167
Time: Coverage Probability	0.9490	0.9360	0.7340	0.9540
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0027	0.0028	0.0056	0.0039
Treatment: SE of Coefficient	0.2702	0.2637	0.2326	0.2780
Treatment: Coverage Probability	0.9530	0.9420	0.8930	0.9560
Treatment: Power	0.9580	0.9620	0.9770	0.9470
Time by Treatment Interaction: Bias	-0.0004	-0.0004	-0.0005	-0.0003
Time by Treatment Interaction: SE of Coefficient	0.0224	0.0212	0.0139	0.0236
Time by Treatment Interaction: Coverage Probability	0.9590	0.9430	0.7570	0.9550
Time by Treatment Interaction: Power	0.9950	0.9980	1.0000	0.9910

Table B14. *Unbalanced discrete data with sample size 350 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0038	0.0037	0.0020	0.0025
Intercept: SE of Coefficient	0.1886	0.1854	0.1594	0.1877
Intercept: Coverage Probability	0.9590	0.9560	0.9140	0.9540
Intercept: Power	1.0000	1.0000	1.0000	0.9990
Time: Bias	-0.0004	-0.0004	-0.0003	-0.0003
Time: SE of Coefficient	0.0166	0.0159	0.0104	0.0162
Time: Coverage Probability	0.9520	0.9480	0.7740	0.9370
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	-0.0086	-0.0085	-0.0070	-0.0072
Treatment: SE of Coefficient	0.2667	0.2621	0.2254	0.2654
Treatment: Coverage Probability	0.9590	0.9570	0.8980	0.9470
Treatment: Power	0.9610	0.9660	0.9790	0.9590
Time by Treatment Interaction: Bias	0.0003	0.0003	0.0002	0.0002
Time by Treatment Interaction: SE of Coefficient	0.0234	0.0225	0.0147	0.0229
Time by Treatment Interaction: Coverage Probability	0.9550	0.9490	0.7770	0.9470
Time by Treatment Interaction: Power	0.9950	0.9970	1.0000	0.9960

Table B15. *Unbalanced discrete data with sample size 350 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0027	-0.0026	-0.0039	-0.0033
Intercept: SE of Coefficient	0.1683	0.1624	0.1372	0.1708
Intercept: Coverage Probability	0.9570	0.9510	0.8830	0.9520
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0000	-0.0001	0.0000	-0.0000
Time: SE of Coefficient	0.0150	0.0139	0.0082	0.0152
Time: Coverage Probability	0.9660	0.9510	0.7480	0.9340
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0044	0.0047	0.0066	0.0063
Treatment: SE of Coefficient	0.2390	0.2306	0.1945	0.2419
Treatment: Coverage Probability	0.9540	0.9490	0.8890	0.9530
Treatment: Power	0.9910	0.9950	0.9960	0.9870
Time by Treatment Interaction: Bias	-0.0007	-0.0008	-0.0009	-0.0009
Time by Treatment Interaction: SE of Coefficient	0.0213	0.0198	0.0118	0.0215
Time by Treatment Interaction: Coverage Probability	0.9630	0.9470	0.7140	0.9500
Time by Treatment Interaction: Power	1.0000	1.0000	1.0000	0.9990

Table B16. *Unbalanced discrete data with sample size 500 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0044	-0.0043	-0.0048	-0.0040
Intercept: SE of Coefficient	0.1605	0.1558	0.1376	0.1649
Intercept: Coverage Probability	0.9530	0.9480	0.9030	0.9570
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0000	-0.0000	0.0000	-0.0001
Time: SE of Coefficient	0.0134	0.0125	0.0082	0.0140
Time: Coverage Probability	0.9430	0.9280	0.7650	0.9480
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	-0.0010	-0.0009	0.0019	0.0006
Treatment: SE of Coefficient	0.2279	0.2212	0.1950	0.2336
Treatment: Coverage Probability	0.9610	0.9520	0.9030	0.9590
Treatment: Power	0.9950	0.9950	0.9970	0.9930
Time by Treatment Interaction: Bias	-0.0002	-0.0002	-0.0005	-0.0004
Time by Treatment Interaction: SE of Coefficient	0.0191	0.0178	0.0117	0.0199
Time by Treatment Interaction: Coverage Probability	0.9600	0.9430	0.7460	0.9570
Time by Treatment Interaction: Power	1.0000	1.0000	1.0000	0.9990

Table B17. *Unbalanced discrete data with sample size 500 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0006	0.0009	0.0004	0.0004
Intercept: SE of Coefficient	0.1573	0.1550	0.1333	0.1574
Intercept: Coverage Probability	0.9510	0.9510	0.8920	0.9500
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0005	-0.0005	-0.0004	-0.0004
Time: SE of Coefficient	0.0137	0.0133	0.0087	0.0136
Time: Coverage Probability	0.9580	0.9540	0.7940	0.9600
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0058	0.0052	0.0037	0.0039
Treatment: SE of Coefficient	0.2234	0.2201	0.1891	0.2231
Treatment: Coverage Probability	0.9560	0.9480	0.9060	0.9470
Treatment: Power	0.9970	0.9970	0.9990	0.9960
Time by Treatment Interaction: Bias	-0.0004	-0.0004	-0.0002	-0.0002
Time by Treatment Interaction: SE of Coefficient	0.0195	0.0189	0.0124	0.0193
Time by Treatment Interaction: Coverage Probability	0.9550	0.9530	0.7760	0.9480
Time by Treatment Interaction: Power	0.9980	0.9980	0.9990	0.9970

Table B18. *Unbalanced discrete data with sample size 500 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0020	0.0026	0.0056	0.0052
Intercept: SE of Coefficient	0.3081	0.2972	0.2499	0.3119
Intercept: Coverage Probability	0.9520	0.9420	0.8790	0.9370
Intercept: Power	0.9150	0.9260	0.9500	0.8980
Time: Bias	0.0002	0.0002	-0.0001	0.0000
Time: SE of Coefficient	0.0270	0.0252	0.0149	0.0275
Time: Coverage Probability	0.9480	0.9290	0.6820	0.9450
Time: Power	0.9600	0.9680	0.9870	0.9400
Treatment: Bias	-0.0053	-0.0062	-0.0088	-0.0085
Treatment: SE of Coefficient	0.4382	0.4227	0.3556	0.4426
Treatment: Coverage Probability	0.9590	0.9530	0.8770	0.9470
Treatment: Power	0.6310	0.6620	0.7570	0.6290
Time by Treatment Interaction: Bias	-0.0012	-0.0011	-0.0007	-0.0008
Time by Treatment Interaction: SE of Coefficient	0.0387	0.0361	0.0216	0.0390
Time by Treatment Interaction: Coverage Probability	0.9550	0.9450	0.7020	0.9480
Time by Treatment Interaction: Power	0.7280	0.7720	0.9340	0.6890

Table B19. *Continuous data with sample size 150 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0056	0.0069	0.0114	0.0105
Intercept: SE of Coefficient	0.2945	0.2849	0.2508	0.3011
Intercept: Coverage Probability	0.9520	0.9420	0.9000	0.9540
Intercept: Power	0.9250	0.9310	0.9520	0.9150
Time: Bias	-0.0011	-0.0012	-0.0015	-0.0014
Time: SE of Coefficient	0.0245	0.0226	0.0149	0.0254
Time: Coverage Probability	0.9520	0.9280	0.7390	0.9500
Time: Power	0.9840	0.9910	0.9980	0.9730
Treatment: Bias	-0.0165	-0.0172	-0.0206	-0.0206
Treatment: SE of Coefficient	0.4189	0.4053	0.3568	0.4275
Treatment: Coverage Probability	0.9570	0.9500	0.9000	0.9520
Treatment: Power	0.6540	0.6710	0.7310	0.6280
Time by Treatment Interaction: Bias	0.0010	0.0011	0.0014	0.0015
Time by Treatment Interaction: SE of Coefficient	0.0351	0.0324	0.0215	0.0361
Time by Treatment Interaction: Coverage Probability	0.9560	0.9310	0.7420	0.9570
Time by Treatment Interaction: Power	0.8290	0.8630	0.9470	0.7910

Table B20. *Continuous data with sample size 150 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0034	-0.0031	-0.0029	-0.0034
Intercept: SE of Coefficient	0.2879	0.2830	0.2426	0.2873
Intercept: Coverage Probability	0.9610	0.9570	0.8950	0.9470
Intercept: Power	0.9350	0.9410	0.9630	0.9340
Time: Bias	-0.0003	-0.0003	-0.0003	-0.0002
Time: SE of Coefficient	0.0249	0.0240	0.0157	0.0245
Time: Coverage Probability	0.9670	0.9610	0.7890	0.9480
Time: Power	0.9860	0.9920	0.9980	0.9830
Treatment: Bias	0.0048	0.0045	0.0024	0.0026
Treatment: SE of Coefficient	0.4097	0.4027	0.3455	0.4082
Treatment: Coverage Probability	0.9590	0.9550	0.8980	0.9470
Treatment: Power	0.6990	0.7120	0.7950	0.6890
Time by Treatment Interaction: Bias	-0.0008	-0.0007	-0.0004	-0.0005
Time by Treatment Interaction: SE of Coefficient	0.0356	0.0344	0.0227	0.0350
Time by Treatment Interaction: Coverage Probability	0.9620	0.9560	0.7820	0.9510
Time by Treatment Interaction: Power	0.8010	0.8230	0.9350	0.8010

Table B21. *Continuous data with sample size 150 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0032	-0.0004	0.0004	0.0007
Intercept: SE of Coefficient	0.1911	0.1963	0.1646	0.2061
Intercept: Coverage Probability	0.9230	0.9540	0.8820	0.9520
Intercept: Power	0.9990	0.9980	0.9980	0.9980
Time: Bias	0.0001	-0.0005	-0.0004	-0.0005
Time: SE of Coefficient	0.0227	0.0167	0.0099	0.0181
Time: Coverage Probability	0.9760	0.9460	0.7090	0.9380
Time: Power	0.9880	1.0000	1.0000	1.0000
Treatment: Bias	0.0049	0.0015	0.0016	0.0011
Treatment: SE of Coefficient	0.2709	0.2783	0.2334	0.2919
Treatment: Coverage Probability	0.9210	0.9510	0.8740	0.9540
Treatment: Power	0.9520	0.9510	0.9720	0.9280
Time by Treatment Interaction: Bias	-0.0006	-0.0000	-0.0002	-0.0001
Time by Treatment Interaction: SE of Coefficient	0.0323	0.0237	0.0140	0.0257
Time by Treatment Interaction: Coverage Probability	0.9840	0.9480	0.7040	0.9500
Time by Treatment Interaction: Power	0.8670	0.9890	0.9990	0.9630

Table B22. *Continuous data with sample size 350 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0050	-0.0048	-0.0052	-0.0046
Intercept: SE of Coefficient	0.1931	0.1877	0.1649	0.1985
Intercept: Coverage Probability	0.9500	0.9440	0.8960	0.9530
Intercept: Power	0.9980	0.9980	0.9990	0.9980
Time: Bias	-0.0001	-0.0002	-0.0001	-0.0002
Time: SE of Coefficient	0.0160	0.0149	0.0098	0.0168
Time: Coverage Probability	0.9440	0.9280	0.7400	0.9520
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0100	0.0100	0.0093	0.0083
Treatment: SE of Coefficient	0.2739	0.2661	0.2339	0.2813
Treatment: Coverage Probability	0.9530	0.9460	0.9010	0.9510
Treatment: Power	0.9630	0.9630	0.9780	0.9520
Time by Treatment Interaction: Bias	-0.0010	-0.0009	-0.0009	-0.0007
Time by Treatment Interaction: SE of Coefficient	0.0227	0.0212	0.0140	0.0237
Time by Treatment Interaction: Coverage Probability	0.9450	0.9260	0.7680	0.9440
Time by Treatment Interaction: Power	0.9870	0.9920	0.9990	0.9860

Table B23. *Continuous data with sample size 350 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0014	-0.0003	-0.0016	-0.0015
Intercept: SE of Coefficient	0.1950	0.1873	0.1601	0.1900
Intercept: Coverage Probability	0.9620	0.9540	0.9110	0.9510
Intercept: Power	0.9990	0.9990	1.0000	1.0000
Time: Bias	-0.0002	-0.0003	-0.0001	-0.0001
Time: SE of Coefficient	0.0180	0.0159	0.0104	0.0163
Time: Coverage Probability	0.9670	0.9540	0.7830	0.9420
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0053	0.0042	0.0030	0.0025
Treatment: SE of Coefficient	0.2766	0.2656	0.2272	0.2692
Treatment: Coverage Probability	0.9530	0.9530	0.9000	0.9480
Treatment: Power	0.9590	0.9690	0.9850	0.9640
Time by Treatment Interaction: Bias	-0.0007	-0.0006	-0.0004	-0.0004
Time by Treatment Interaction: SE of Coefficient	0.0256	0.0226	0.0148	0.0231
Time by Treatment Interaction: Coverage Probability	0.9680	0.9460	0.7920	0.9480
Time by Treatment Interaction: Power	0.9830	0.9940	0.9970	0.9870

Table B24. *Continuous data with sample size 350 and Linear correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	0.0048	0.0047	0.0042	0.0047
Intercept: SE of Coefficient	0.1700	0.1640	0.1378	0.1717
Intercept: Coverage Probability	0.9540	0.9460	0.8840	0.9490
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0007	-0.0007	-0.0007	-0.0008
Time: SE of Coefficient	0.0150	0.0140	0.0083	0.0152
Time: Coverage Probability	0.9600	0.9480	0.7400	0.9400
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0026	0.0026	0.0033	0.0023
Treatment: SE of Coefficient	0.2403	0.2318	0.1952	0.2430
Treatment: Coverage Probability	0.9530	0.9430	0.8930	0.9480
Treatment: Power	0.9850	0.9880	0.9930	0.9830
Time by Treatment Interaction: Bias	-0.0005	-0.0004	-0.0005	-0.0004
Time by Treatment Interaction: SE of Coefficient	0.0212	0.0198	0.0117	0.0215
Time by Treatment Interaction: Coverage Probability	0.9630	0.9480	0.7320	0.9510
Time by Treatment Interaction: Power	0.9980	0.9980	1.0000	0.9970

Table B25. *Continuous data with sample size 500 and Gaussian correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0011	-0.0009	-0.0026	-0.0014
Intercept: SE of Coefficient	0.1586	0.1571	0.1382	0.1659
Intercept: Coverage Probability	0.9560	0.9560	0.9080	0.9570
Intercept: Power	1.0000	1.0000	1.0000	0.9990
Time: Bias	-0.0003	-0.0003	-0.0001	-0.0003
Time: SE of Coefficient	0.0128	0.0125	0.0082	0.0140
Time: Coverage Probability	0.9350	0.9350	0.7570	0.9470
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	-0.0001	-0.0003	0.0021	0.0009
Treatment: SE of Coefficient	0.2243	0.2221	0.1959	0.2349
Treatment: Coverage Probability	0.9520	0.9500	0.9060	0.9550
Treatment: Power	0.9950	0.9950	0.9960	0.9930
Time by Treatment Interaction: Bias	-0.0006	-0.0006	-0.0010	-0.0008
Time by Treatment Interaction: SE of Coefficient	0.0181	0.0177	0.0117	0.0199
Time by Treatment Interaction: Coverage Probability	0.9420	0.9350	0.7550	0.9620
Time by Treatment Interaction: Power	1.0000	1.0000	1.0000	1.0000

Table B26. *Continuous data with sample size 500 and Exponential correlation*

	corExp	corGaus	HLM1	HLM2
Intercept: Bias	-0.0013	-0.0011	-0.0015	-0.0007
Intercept: SE of Coefficient	0.1583	0.1565	0.1341	0.1585
Intercept: Coverage Probability	0.9470	0.9460	0.8880	0.9380
Intercept: Power	1.0000	1.0000	1.0000	1.0000
Time: Bias	-0.0002	-0.0003	-0.0002	-0.0003
Time: SE of Coefficient	0.0136	0.0133	0.0087	0.0136
Time: Coverage Probability	0.9630	0.9560	0.7880	0.9510
Time: Power	1.0000	1.0000	1.0000	1.0000
Treatment: Bias	0.0014	0.0011	-0.0003	-0.0011
Treatment: SE of Coefficient	0.2239	0.2214	0.1901	0.2244
Treatment: Coverage Probability	0.9560	0.9520	0.9020	0.9440
Treatment: Power	0.9950	0.9950	0.9970	0.9920
Time by Treatment Interaction: Bias	-0.0001	-0.0001	0.0000	0.0001
Time by Treatment Interaction: SE of Coefficient	0.0193	0.0189	0.0123	0.0192
Time by Treatment Interaction: Coverage Probability	0.9540	0.9520	0.7810	0.9380
Time by Treatment Interaction: Power	1.0000	1.0000	1.0000	1.0000

Table B27. *Continuous data with sample size 500 and Linear correlation*

Appendix C

Derivation of the Covariance Structure for the GSC and HLM2

For the GSC model, consistent with our model specification in the The General Serial Covariance Model section, the derivation can be written as follows:

$$\begin{aligned} var(Y_{ij}) &= var(\mu_{ij} + \alpha_i + w_i(t_{ij}) + \varepsilon_{ij}) \\ &= var(\alpha_i + w_i(t_{ij}) + \varepsilon_{ij}) \\ &= \nu^2 + \tau^2 + \sigma^2 \end{aligned} \tag{6}$$

and for $j \neq k$,

$$\begin{aligned} cov(Y_{ij}, Y_{ik}) &= cov(\mu_{ij} + \alpha_i + w_i(t_{ij}) + \varepsilon_{ij}, \mu_{ik} + \alpha_i + w_i(t_{ik}) + \varepsilon_{ik}) \\ &= cov(\alpha_i, \alpha_i) + cov(\alpha_i, w_i(t_{ik})) + cov(\alpha_i, \varepsilon_{ik}) + \\ &\quad cov(w_i(t_{ij}), \alpha_i) + cov(w_i(t_{ij}), w_i(t_{ik}) + cov(w_i(t_{ij}), \varepsilon_{ik}) + \\ &\quad cov(\varepsilon_{ij}, \alpha_i) + cov(\varepsilon_{ij}, w_i(t_{ik})) + cov(\varepsilon_{ij}, \varepsilon_{ik}) + \\ &= \nu^2 + \tau^2 \rho(t_{ij} - t_{ik}), \end{aligned} \tag{7}$$

where the function ρ can be any spatial covariance function such as Exponential, Gaussian, or Linear, as defined in the text.

Additionally, let $Y_{ij} = \mu_{ij} + \theta_{i1} + \theta_{i2} t_{ij} + \varepsilon_{ij}$ where μ_{ij} is the fixed effect similar to the fixed effect in the GSC model defined in-text. The HLM2 model assumptions are as follows:

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim \text{MVN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right] \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2)$$

where θ_{i1} and θ_{i2} are independent of ε_{ij}. Additionally, the ε_{ij}s are independent for all the repeated measures (i.e. for all js). Then a random intercept and slope model (i.e. HLM2) induces the following variance/covariance structure:

$$\begin{aligned} var(Y_{ij}) &= var(\mu_{ij} + \theta_{i1} + \theta_{i2} t_{ij} + \varepsilon_{ij}) \\ &= var(\theta_{i1} + \theta_{i2} t_{ij} + \varepsilon_{ij}) \\ &= var(\theta_{i1}) + var(\theta_{i2} t_{ij}) + var(\varepsilon_{ij}) + \\ &\quad 2cov(\theta_{i1}, \theta_{i2} t_{ij}) + 2cov(\theta_{i1}, \varepsilon_{ij}) + 2cov(\theta_{i2} t_{ij}, \varepsilon_{ij}) \\ &= \sigma_1^2 + \sigma_2^2 t_{ij}^2 + \sigma^2 + 2\sigma_{12} t_{ij} \\ &= \sigma_1^2 + \sigma_2^2 t_{ij}^2 + \sigma^2 + 2\sigma_{12} t_{ij} \end{aligned} \tag{8}$$

For $j \neq k$,

$$
\begin{aligned}
cov(Y_{ij}, Y_{ik}) &= cov(\theta_{i1} + \theta_{i2}t_{ij} + \varepsilon_{ij}, \theta_{i1} + \theta_{i2}t_{ik} + \varepsilon_{ik}) \\
&= cov(\theta_{i1}, \theta_{i1}) + cov(\theta_{i1}, \theta_{i2}t_{ik}) + cov(\theta_{i1}, \varepsilon_{ik}) + \\
&\quad cov(\theta_{i2}t_{ij}, \theta_{i1}) + cov(\theta_{i2}t_{ij}, \theta_{i2}t_{ik}) + cov(\theta_{i2}t_{ij}, t_{ik}, \varepsilon_{ik}) + \\
&\quad cov(\varepsilon_{ij}, \theta_{i1}) + cov(\varepsilon_{ij}, \theta_{i2}t_{ik}) + cov(\varepsilon_{ij}, \varepsilon_{ik}) \\
&= \sigma_1^2 + \sigma_{12}t_{ik} + \sigma_{12}t_{ij} + \sigma_2^2 t_{ij}t_{ik} \\
&= \sigma_1^2 + \sigma_{12}(t_{ik} + t_{ij}) + \sigma_2^2 t_{ij}t_{ik}.
\end{aligned}
\tag{9}
$$

Taking the derivative of the $var(Y_{ij})$ with respect to time results in:

$$
\frac{\partial var(Y_{ij})}{\partial t_{ij}} = 2\sigma_2^2 t_{ij} + 2\sigma_{12} \quad
\begin{cases}
> 0 & \text{if } t_{ij} > -\frac{\sigma_{12}}{\sigma_2^2} \\
= 0 & \text{if } t_{ij} = -\frac{\sigma_{12}}{\sigma_2^2} \\
< 0 & \text{if } t_{ij} < -\frac{\sigma_{12}}{\sigma_2^2}.
\end{cases}
\tag{10}
$$

From (10), one can observe the following:

- When σ_{12} is positive then $\frac{\partial var(Y_{ij})}{\partial t_{ij}}$ indicates that $var(Y_{ij})$ is an increasing function of time.
- When σ_{12} is negative then $var(Y_{ij})$ is first a decreasing function of time then as time increases $var(Y_{ij})$ will eventually be an increasing function of time.

Moreover, taking the partial derivative of the $cov(Y_{ij}, Y_{ik})$ with respect to time (i.e. t_{ij} and t_{ik} separately) results in:

$$
\frac{\partial cov(Y_{ij}, Y_{ik})}{\partial t_{ij}} = \sigma_{12} + \sigma_{12}^2 t_{ik} \quad
\begin{cases}
> 0 & \text{if } t_{ik} > -\frac{\sigma_{12}}{\sigma_2^2} \\
= 0 & \text{if } t_{ik} = -\frac{\sigma_{12}}{\sigma_2^2} \\
< 0 & \text{if } t_{ik} < -\frac{\sigma_{12}}{\sigma_2^2}.
\end{cases}
\tag{11}
$$

Looking at (11), a similar argument as (10) implies here. Taking the partial derivative with respect to t_{ik} is similar to above derivations and results.

Appendix D
R Code

```
#---------------------------------------------------------------------------
# Functions to be used for the GSC simulations and estimations
#---------------------------------------------------------------------------

# Extracting parameters for GSC ------------------------------------
extract_lme = function(fit) {
  nugget = coef(fit$modelStruct$corStruct, unconstrained = F)[2]
  residual = fit$sigma

  sigma_hat = sqrt(residual^2*nugget)
  tau_hat = sqrt(residual^2 - sigma_hat^2)
  phi_hat = coef(fit$modelStruct$corStruct, unconstrained = F)[1]
  nu_hat = as.numeric(VarCorr(fit)[1, 2])


  value = c(as.numeric(summary(fit)$tTable[, 1]),
            as.numeric(summary(fit)$tTable[, 2]),
            nu_hat, sigma_hat, tau_hat, phi_hat)


  names(value) = c(rownames(summary(fit)$tTable), paste0(rownames(summary(fit)
      $tTable), "_se"),
                 "nu", "sigma", "tau", "phi")
  return(value)
}

# Extracting parameters for HLM -----------------------------------
extract_HLM = function(fit) {

  value = c(as.numeric(summary(fit)$tTable[, 1]),
            as.numeric(summary(fit)$tTable[, 2]))

  names(value) = c(rownames(summary(fit)$tTable), paste0(rownames(summary(fit)
      $tTable), "_se"))
  return(value)
}

# find the index for each person------------------------------------
# return a list, with each element being the indexes for the person
find_index_for_person = function(ni) {
  n = length(ni)
  cumsum_ni = cumsum(ni)
  index_for_person = list()
  for (i in 1:n) {
    if (i==1) {
```

```
      index_for_person[[i]] = 1:cumsum_ni[i]
    } else {
      index_for_person[[i]] = (cumsum_ni[i-1] + 1):cumsum_ni[i]
    }
  }
  return(index_for_person)
}

#-------------------------------------------------------------------------
# Simulating data from the GSC model which has
# random intercept, serial correlation and measurement error
# sim_SIG function specifications are as follows:
# n = number of individuals
# ni = the number of observations of each person
# time = collection of all the time indices - it's time but not the index per
    person!
# X = design matrix
# beta = coefficent vector
# nu = SD of random intercept
# sigma = SD of measurment error
# cor_str = type of serial correlation, can be "linear", "spatial", "gaussian", "
    exponential"
# tau = SD component corresponding to the serial correlation
# phi = rate of decay, used for gaussian and exponential corr structure (look at
    formulas in the next function)
# d = used with linear correlation structure (look at formulas in the next function
    )

sim_GSC = function(n, ni, time, X, beta, nu, sigma, cor_str, tau, phi = NA, d = NA)
     {
  index_for_person = find_index_for_person(ni)

# Covariance matrix of the serial correlated noise
  V = list()
  for (i in 1:n) {
    tk <- matrix(time[index_for_person[[i]]], ncol = ni[i], nrow = ni[i], byrow = F
        )
    tj <- matrix(time[index_for_person[[i]]], ncol = ni[i], nrow = ni[i], byrow = T
        )
    #using R parametrization here written in blue notebook under extracting
        parameters in R
    V[[i]] = tau^2*switch(cor_str,
                          exponential = exp(-abs(tk-tj)/phi),
                          gaussian = exp(-(tk-tj)^2/phi^2),
                          linear = (1-abs(tk-tj)/d)*(abs(tk-tj) <d))
  }
```

```
  # Simulation
  # beta is the regression parameters for the covariates
  Y = X%*%beta # first assign the mean (X%*%beta) to the Y
  for (i in 1:n) {
    alpha_i = rnorm(1, 0, sd = nu) #random intercept
    epsilon = rnorm(ni[i], 0, sd = sigma) # measurment error
    Y[index_for_person[[i]]] = Y[index_for_person[[i]]] + alpha_i + epsilon +
        mvrnorm(1, mu = rep(0, ni[i]), Sigma = V[[i]])
  }

  data = data.frame(indv = rep(1:n, ni), time = time, Y = Y, treatment = X[, 3],
      time_treatment = X[, 4])
  return(data)
}

# Creating X matrix for balance and imbalance data--------------
sim_X <- function(n, balance, cts){
  set.seed(9292018)
  if (balance) {
    ni = rep(15, n) #balance design
  } else {
    # no. of observations
    ni = sample(10:15, n, replace = T) # unbalance design
  }

  X = matrix(0, nrow = sum(ni), ncol = 4) # design matrix
  time = rep(0, sum(ni)) # vector of length sum^n_i=1(ni)
  index_for_person = find_index_for_person(ni)
  # balance=TRUE
  for (i in 1:n) {
    if (cts) {
      # uniform missing from 1:15, no. of observations = ni[i]
      time[index_for_person[[i]]] = sort(sample(1:15, ni[i], replace = FALSE)) +
          runif(ni[i],0, 0.25)
    } else {
      time[index_for_person[[i]]] = sort(sample(1:15, ni[i], replace = FALSE))
    }
    X[index_for_person[[i]], 1] = 1 # intercept
    X[index_for_person[[i]], 2] = time[index_for_person[[i]]] # time
    if (i <= n/2) {
      X[index_for_person[[i]], 3] = 1 #treatment 0/1
    }
  }
  X[, 4] = X[,2]*X[,3] #interaction between time and treatment

  colnames(X)=c("Intercept","time","Treatment","TimeXTreatment")
```

```
  output = list()
  output[[1]] = X
  output[[2]] = time
  output[[3]] = ni
  names(output) = c("X", "time", "ni")
  return(output)
}

find_bias_se_covg = function(n, ni, time, X, beta, nu, sigma,
                              cor_str, tau, phi = NA, d = NA, no_sim, save_data =
                                 FALSE, save_data_name = NA) {
  name_str = c("corExp", "corGaus","HLM1","HLM2")
 #HLM1=Random intercept, HLM2=Random intercept and random slope

  estimate = list()
  estimate[[1]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 12))
  estimate[[2]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 12))
  estimate[[3]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 8))
  estimate[[4]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 8))

  names(estimate) = name_str
  colnames(estimate[[1]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3", "nu", "sigma", "tau", "phi")
  colnames(estimate[[2]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3", "nu", "sigma", "tau", "phi")
  colnames(estimate[[3]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3" )
  colnames(estimate[[4]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3")

  choose_right_model = rep(0, no_sim*10)
  problematic_seed_model = matrix(1, no_sim*10, length(name_str))
  colnames(problematic_seed_model) = c("Exp", "Gaus", "HLM1", "HLM2")

  i=0
  common_index=0
  while(length(common_index)<no_sim){
    i=i+1
    set.seed(i)
    data =sim_GSC(n, ni, time, X, beta, nu, sigma, cor_str, tau, phi, d)
    if (save_data) {
      write.csv(data$Y, paste0(save_data_name, "_", i, ".csv"))
    }

    tryCatch({
```

```
  fit = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
             correlation = corExp( form = ~ time| indv, nugget=TRUE),
             data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 1] = 0 # meaning no problem

  # print(paste(i, "corExp is finished"))
  estimate$corExp[i, 1:(dim(estimate[[1]])[2])] = extract_lme(fit)
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with corExp() does not converge", sep ="")
      )
})

tryCatch({
  fit = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
             correlation = corGaus( form = ~ time| indv, nugget=TRUE),
             data = data, control=lmeControl(opt="optim"))

  problematic_seed_model[i, 2] = 0 # meaning no problem
  # print(paste(i, "corGaus is finished"))
  estimate$corGaus[i, 1:(dim(estimate[[1]])[2])] = extract_lme(fit)
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with corGaus() does not converge", sep
      =""))
})


tryCatch({
  HLM1 = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
             data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 3] = 0 # meaning no problem
  estimate$HLM1[i, ] = c(extract_HLM(HLM1))
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with HLM with only random intercept does
      not converge", sep =""))
})

tryCatch({
  HLM2 = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 +time | indv, pdClass="pdSymm"),
             data = data, control=lmeControl(opt="optim"))

  problematic_seed_model[i, 4] = 0 # meaning no problem

  estimate$HLM2[i, ] = c(extract_HLM(HLM2))
```

```
  }, error = function(e) {
    print(paste("Seed ", i, ": Fitting with HLM with random intercept & slope
        does not converge", sep =""))
  })

  print(paste("Trial: ", i, sep =""))

  # Determining when to stop the while loop
  can_estimate = list()
  for (j in 1:length(name_str)) {
    #index of the times that can be estimated is equivqlent to the index of the
        non-zero intercepts
    can_estimate[[j]] = which(estimate[[j]][, 1]!=0)
  }

  common_index = Reduce(intersect, can_estimate)

  print(length(common_index))

} # End of while

# This part is to find common index after the loop
can_estimate = list()
for (j in 1:length(name_str)) {
  can_estimate[[j]] = which(estimate[[j]][, 1]!=0)
}

common_index = Reduce(intersect, can_estimate)
for (j in 1:length(name_str)) {
  estimate[[j]] = estimate[[j]][common_index, ]
}
choose_right_model = choose_right_model[common_index]

#---------------------------------------------------------------------------
bias = matrix(0, nrow = length(name_str), ncol= length(beta))
se = matrix(0, nrow = length(name_str), ncol = length(beta))
est_sd = matrix(0, nrow = length(name_str), ncol = length(beta))

IC = matrix(0, nrow = length(name_str), ncol = 2)
colnames(bias) = c("Beta0", "Beta1", "Beta2", "Beta3")
colnames(se) = c("SE.Beta0", "SE.Beta1", "SE.Beta2", "SE.Beta3")
colnames(est_sd) = c("SD.Beta0", "SD.Beta1", "SD.Beta2", "SD.Beta3")

rownames(bias) = name_str
rownames(se) = name_str
rownames(est_sd) = name_str
```

```
  rownames(IC) = name_str

  for (j in 1:length(name_str)) {
    bias[j, ] = colMeans(estimate[[j]][, 1:length(beta)])-beta
    se[j, ] = colMeans(estimate[[j]][, (1:length(beta))+ length(beta) ]) # this is
        just col mean of all SEs for each parameter
    est_sd[j, ] = apply(estimate[[j]][, 1:length(beta)], 2, sd) # this is SD of
        estimated parametrs beta0 to beta3
  }

# Calculating Coverage Probability
  coverage = list()
  for (m in 1:length(name_str)) {
    coverage[[m]] = matrix(0, nrow = no_sim, ncol = length(beta))
    colnames(coverage[[m]]) = c("coverage.Beta0", "coverage.Beta1", "coverage.Beta2
        ", "coverage.Beta3")
    for (j in 1:no_sim) {
      for (k in 1:length(beta)) {
        L = estimate[[m]][j, k] - abs(qnorm(0.025))*estimate[[m]][j, k+4]
        U = estimate[[m]][j, k] + abs(qnorm(0.025))*estimate[[m]][j, k+4]
        coverage[[m]][j, k] = as.numeric(beta[k] > L & beta[k] < U)
      }
    }

  }

  names(coverage)= name_str

  mean_coverage = matrix(0, nrow = length(name_str), ncol =4)
  rownames(mean_coverage) = name_str
  colnames(mean_coverage) = c("coverage.Beta0", "coverage.Beta1", "coverage.Beta2",
      "coverage.Beta3")

  for (s in 1:length(name_str)) {
    mean_coverage[s, ] = colMeans(coverage[[s]])
  }

#---------------------------------------------------------------------------

  value = list()
  value[[1]] = estimate
  value[[2]] = bias
  value[[3]] = se
  value[[4]] = problematic_seed_model[1:i, ]
  value[[5]] = i
  value[[6]] = coverage
  value[[7]] = mean_coverage
```

```
  value[[8]] = est_sd

  names(value) = c("estimate", "bias", "se", "problematic_seed_model",
                 "number_trial", "coverage", "mean_coverage", "est_sd")
  return(value)
}

#Function for finding power of two tail test -----------------------
find_power = function(n, ni, time, X, beta, beta1, nu, sigma, cor_str, tau, phi =
    NA, d = NA,
                      no_sim, save_data = FALSE, save_data_name = NA) {

  name_str = c("corExp", "corGaus", "HLM1", "HLM2")
  #HLM1=Random intercept, HLM2=Random intercept and random slope

  estimate = list()
  estimate[[1]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 12))
  estimate[[2]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 12))
  estimate[[3]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 8))
  estimate[[4]] = data.frame(matrix(0, nrow = no_sim*10, ncol = 8))


  names(estimate) = name_str
  colnames(estimate[[1]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3", "nu", "sigma", "tau", "phi")
  colnames(estimate[[2]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3", "nu", "sigma", "tau", "phi")
  colnames(estimate[[3]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3")
  colnames(estimate[[4]]) = c("Beta0", "Beta1", "Beta2", "Beta3","SE.Beta0", "SE.
     Beta1", "SE.Beta2", "SE.Beta3")

  choose_right_model = rep(0, no_sim*10)
  problematic_seed_model = matrix(1, no_sim*10, length(name_str))
  colnames(problematic_seed_model) = c("Exp", "Gaus", "HLM1", "HLM2")

  i=0
  common_index = 0
  while(length(common_index) < no_sim){
    i=i+1
    set.seed(i)
    data =sim_GSC(n, ni, time, X, beta1, nu, sigma, cor_str, tau, phi, d )
    if (save_data) {
      write.csv(data$Y, paste0(save_data_name, "_", i, ".csv"))
    }

    tryCatch({
```

```
  fit = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
              correlation = corExp( form = ~ time| indv, nugget=TRUE),
              data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 1] = 0 # meaning no problem
  # print(paste(i, "corExp is finished"))
  estimate$corExp[i, 1:(dim(estimate[[1]])[2])] = extract_lme(fit)
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with corExp() does not converge", sep ="")
      )
})

tryCatch({
  fit = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
              correlation = corGaus( form = ~ time| indv, nugget=TRUE),
              data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 2] = 0 # meaning no problem
  # print(paste(i, "corGaus is finished"))
  estimate$corGaus[i, 1:(dim(estimate[[1]])[2])] = extract_lme(fit)
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with corGaus() does not converge", sep
      =""))
})

tryCatch({
  HLM1 = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 | indv, pdClass="pdSymm"),
              data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 3] = 0 # meaning no problem
  estimate$HLM1[i,] = c(extract_HLM(HLM1))
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with HLM with only random intercept does
      not converge", sep =""))
})

tryCatch({
  HLM2 = lme( Y ~ time + treatment + time_treatment, method = "ML", random =
      reStruct( ~ 1 +time | indv, pdClass="pdSymm"),
              data = data, control=lmeControl(opt="optim"))
  problematic_seed_model[i, 4] = 0 # meaning no problem
  estimate$HLM2[i,] = c(extract_HLM(HLM2))
}, error = function(e) {
  print(paste("Seed ", i, ": Fitting with HLM with random intercept & slope
      does not converge", sep =""))
})
```

```
    print(paste("Trial: ", i, sep =""))

    can_estimate = list()
    #index of the times that can be estimated is equivalent to the index of the non
       -zero intercepts
    for (j in 1:length(name_str)) {
      can_estimate[[j]] = which(estimate[[j]][, 1]!=0) # using only intercept non-
         zero
    }

    common_index = Reduce(intersect, can_estimate) # the index for which all the
       models can be estimated

    print(length(common_index))

  } # End of while


  can_estimate = list()
  for (j in 1:length(name_str)) {
    can_estimate[[j]] = which(estimate[[j]][, 1]!=0)
  }

  common_index = Reduce(intersect, can_estimate)
  for (j in 1:length(name_str)) {
    estimate[[j]] = estimate[[j]][common_index, ]
  }
  choose_right_model = choose_right_model[common_index]


## Calculating power-----------------------------------------------------
## Data is simulated under alternative hypothesis
  power = list()
  for (m in 1:length(name_str)) {
    power[[m]] = matrix(0, nrow = no_sim, ncol = length(beta1))
    colnames(power[[m]]) = c("power.Beta0", "power.Beta1", "power.Beta2", "power.
       Beta3")
    for (j in 1:no_sim) {
      for (k in 1:length(beta1)) {
        L = estimate[[m]][j, k] - abs(qnorm(0.025))*estimate[[m]][j, k+4]
        U = estimate[[m]][j, k] + abs(qnorm(0.025))*estimate[[m]][j, k+4]
        power[[m]][j, k] = as.numeric(beta[k] < L | beta[k] > U)
      }
    }
  }
  names(power)= name_str
```

```
  mean_power = matrix(0, nrow = length(name_str), ncol =4)
  rownames(mean_power) = name_str
  colnames(mean_power) = c("power.Beta0", "power.Beta1", "power.Beta2", "power.
      Beta3")

  for (s in 1:length(name_str)) {
    mean_power[s, ] = colMeans(power[[s]])
  }

  #-------------------------------------------------------------------------
  value = list()
  value[[1]] = estimate
  value[[2]] = problematic_seed_model[1:i, ]
  value[[3]] = i
  value[[4]] = power
  value[[5]] = mean_power

  names(value) = c("estimate", "problematic_seed_model",
                   "number_trial","power", "mean_power")
  return(value)
}


#-------------------------------------------------------------------------
# To run the following on cluster computing
#-------------------------------------------------------------------------


# Estimation of bias, SE, convergence and power------------------
# Source code for functions above is needed to run this code

rm(list=ls())
arg = Sys.getenv("SLURM_ARRAY_TASK_ID")
arg = as.numeric(arg)
uni = "place holder"
.libPaths(paste("/rigel/home/",uni,"/packages2",sep=""))
library(MASS)
library(nlme)

source(paste0("/rigel/home/", uni, "/GSC/program/functions.R"))


setting_matrix = matrix(0, nrow = 18, ncol = 4)
colnames(setting_matrix) = c("data_setting", "sample_size", "data_cor_str", "
    cp_or_power")
setting_matrix[, 1] = "bal_dis"
setting_matrix[1:6, 2] = 150
```

```
setting_matrix[7:12, 2] = 350
setting_matrix[13:18, 2] = 500
setting_matrix[, 3] = rep(c("gaussian", "exponential", "linear"), 6)
setting_matrix[, 4] = rep(c("cp", "cp", "cp", "power", "power", "power"), 3)
setting_matrix = rbind(setting_matrix, setting_matrix, setting_matrix)
setting_matrix[19:36, 1] = "unbal_dis"
setting_matrix[37:54, 1] = "cts"

data_setting = as.character(setting_matrix[arg, 1])
n = as.numeric(setting_matrix[arg, 2])
cor_str = as.character(setting_matrix[arg, 3])
cp_or_power = as.character(setting_matrix[arg, 4])


# source(".../functions.R")

beta = c(10, 0.5, 6, 2)
beta1 = c(11, 0.6 , 7, 2.1)


nu = 1.5
sigma = 1
tau = 2
phi = 3
d = 3.5
no_sim = 1000


## Extracting the seed -----------------------------------------------
#rownames(unbal_dis_power_lin$estimate$HLM1)


if (data_setting == "bal_dis") {
  temp = sim_X(n, balance = TRUE, cts = FALSE)
}


if (data_setting == "unbal_dis") {
  temp = sim_X(n, balance = FALSE, cts = FALSE)
}

if (data_setting == "cts") {
  temp = sim_X(n, balance = FALSE, cts = TRUE)
}

X = temp$X
ni = temp$ni
```

```r
time = temp$time


dir = paste0("/rigel/home/", uni, "/GSC/", data_setting, "/n", n, "/", cor_str,
    "/")
name = paste0(data_setting, "_n", n, "_", cor_str, "_", cp_or_power)

if (cp_or_power == "cp") {
  summary_fit = find_bias_se_covg(n, ni = ni, time = time, X = X, beta, nu, sigma,
      cor_str, tau, phi , d , no_sim, save_data = FALSE,
                 save_data_name = paste0(dir, cp_or_power, "_data/data"))


  write.csv(summary_fit$bias, paste0(dir, name, "_bias.csv"))
  write.csv(summary_fit$se, paste0(dir, name, "_se.csv"))
  write.csv(summary_fit$est_sd, paste0(dir, name, "_est_sd.csv"))
  write.csv(summary_fit$mean_coverage, paste0(dir, name, "_mean_coverage.csv"))
}

if (cp_or_power == "power") {
  summary_fit = find_power(n, ni = ni, time = time, X = X, beta, beta1, nu, sigma,
      cor_str, tau, phi , d , no_sim, save_data = FALSE,
               save_data_name = paste0(dir, cp_or_power, "_data/data"))

  write.csv(summary_fit$mean_power, paste0(dir, name, "_mean_power.csv"))
}

write.csv(summary_fit$problematic_seed_model, paste0(dir, name, "_problematic_seed.
    csv"))

write.csv(summary_fit$estimate$corExp, paste0(dir, name, "_est_corExp.csv"))
write.csv(summary_fit$estimate$corGaus, paste0(dir, name, "_est_corGaus.csv"))
write.csv(summary_fit$estimate$HLM1, paste0(dir, name, "_est_HLM1.csv"))
write.csv(summary_fit$estimate$HLM2, paste0(dir, name, "_est_HLM2.csv"))
write.csv(as.numeric(rownames(summary_fit$estimate$HLM1)), paste0(dir, name, "_seed
    .csv"))
```

Paper 3: A Tutorial on Using Linear Mix Modeling and Spatial Correlation with an Online Drug Abuse Prevention Intervention Data in R

Hedyeh Ahmadi

Teachers College, Columbia University

ABSTRACT

Paper 3: A Tutorial on Using Linear Mix Modeling and Spatial Correlation with an Online Drug Abuse Prevention Intervention Data in R

Hedyeh Ahmadi

Modeling the correlation structure in repeated measure data is essential for proper data analysis. A survey of longitudinal methods in the first paper showed that this correlation structure is not being modeled optimally in Education and Psychology literature. A simulation study in the second paper showed that when data are consistent with a General Serial Covariance (GSC) model with different spatial correlations, using basic random intercept or random intercept/slope models does not produce optimal estimation and testing properties. A drug abuse prevention intervention data set was analyzed in detail using R programming language. This tutorial first offers a concise exploratory data analysis (EDA) using various tables and plots. As a part of the EDA for longitudinal data, the use of variogram plots is introduced to identify the functional form and different variability components of the covariance structure of the repeated measure. The paper then discusses model fitting and model comparison. Finally, the fixed effect of the GSC model is presented using splines.

Table of Contents

Introduction

Modeling the covariance structure of repeated measure data is critical for longitudinal data analysis because it increases the precision of estimates and improves testing properties. Although the modeling of longitudinal data is prevalent in Education and Psychology, a survey of literature in these disciplines (presented in the first paper) demonstrated that researchers in these fields frequently omit exploring, reporting, and modeling the correlation pattern of the repeated measure data. This means scholars are either not exploring the covariance structure or they are simply using the program defaults. Models such as repeated measure ANOVA and basic Hierarchical Linear Models (HLM) are widely used without any exploratory analysis of the covariance structure of the repeated measure.

Methods such as the General Serial Covariance (GSC) can be used to model the covariance pattern and fixed effects at the same time; the GSC model can be thought of as an HLM model that incorporates the appropriate covariance structure. This model can answer questions such as:

- What is the usual time course of seeing the desired result (such as increased test score or decreased refusal score) after an intervention?
- What are the factors predicting the outcome of interest?
- What are the characteristics of heterogeneity within and across subjects in terms of the outcome of interest?

A detailed introduction to the GSC model was presented in the second paper. A simulation study, also in the second paper, confirmed that running basic HLMs for data consistent with the GSC model with spatial correlation can have a negative effect on the power and the standard error (SE) of the estimates. More specifically, this simulation study showed that when data are consistent with the GSC model with spatial correlation (i.e. Linear, Exponential, and Gaussian), a random intercept-only model has an SE furthest from the true Monte Carlo SE. The study further showed that the random intercept/slope model has the lowest power compared to running a GSC model with Exponential and Gaussian covariance structure.

Although there are many books on longitudinal data analysis in Education and Psychology, simple and easy-to-follow tutorials (using R programming language) on the GSC model that explore the specifications of the repeated measure correlation structure are non-existent. The focus of this tutorial is to teach users how to perform exploratory data analysis (EDA) for longitudinal data by introducing count tables, spaghetti plots, cross-sectional smoothers of the residuals, and variograms. Special attention is given to plotting and interpreting the variograms in the context of the GSC model to explore the covariance pattern of the repeated measure for modeling purposes. HLM and GSC model fitting and model comparison are also presented using plots and tables along with the necessary R functions and command snapshots. Additionally, a short description of splines as a more flexible way of modeling the fixed-effect part of the GSC model is offered. All the necessary R codes for using and presenting the results from Basis Splines and Natural Splines are also provided. Assumption checking and model diagnostics are very important steps that are reserved for future papers.

Data Description

A nationwide (48 states and the District of Columbia) longitudinal online drug abuse prevention program recruited 797 adolescent girls (13 to 14 years old) through Facebook advertisements. Girls who enrolled in the program were randomly assigned to the intervention or control group. All the girls completed the pretest forms online. After nine sessions of the gender-specific drug abuse prevention web-based program, the intervention group was assigned to complete the post-test measures. The control group completed the post-test measures 14 weeks after their pre-test date. Finally, all of the girls completed three follow-up measures, each about 1 year apart. For more details on the data collection, sample characteristics, and Facebook advertisement process, see Schwinn, Hopkins, Schinke, and Liu (2017).

Variable Descriptions

The end time and date of each measurement were converted to one continuous scale and used as the time variable. Note that the end time and date were used in all of the measurements in order for all of the girls to have finished the treatment. Using the R built-in function `as.POSIXct()`, the variable date/time were converted to the number of seconds from January 1st, 1970. Then the minimum of the time variable was subtracted from the continuous time variable to yield the final version of the continuous time variable.

Each wave of the data was recorded separately. The treatment arm variable was merged from the different waves to reduce the missingness for the treatment variable for each individual.

Missing observations were deleted, but all of the individuals were kept in the analysis. There was only one individual with two rows of data (in one of the rows she switched between treatment and control). This was an obvious data recording mistake so it was deleted.

All of the R libraries (R Core Team, 2017) and functions used to produce this tutorial are shown in Appendix A. The data cleaning process for longitudinal data is not shown in this tutorial. Readers should be cautioned that the data was used in either wide *or* long format, depending on the command requirements in R.

Three refusal skill variables (i.e. cigarette, alcohol, and marijuana refusal skills) were used to create a composite refusal variable score. Table 1 represents the description of the variables used in this tutorial. Table B1 in Appendix B shows the coding scheme and the questionnaire for the mean cigarette, alcohol, and marijuana refusal skills.

Variable Name	Variable Description
ID	Participant ID number.
time	The observation wave number.
TRT	Treatment: 0 = Control and 1 = Intervention.
CTS.TIMEDATE	Continuous time variable (in year) in which each participants were measured.
MEAN.CIG.ALC.POT	Total mean refusal skill.

Table 1. *Variable descriptions*

Table 2 can be used to look at all the variable summaries. There are five waves of measurements under the time wave column. The CTS.TIMEDATE variable ranged from 0 to 3.8; mean and median for this variable are around 1.5. There were 1900 observations for the treatment group versus 1829 observations for the control. Finally, the MEAN.CIG.ALC.POT variable ranged from 0 to 5, with mean and median close to 1.8. Note that total mean refusal skill was reverse coded so the lower score means higher refusal skill for cigarette, alcohol, and marijuana.

	ID	time	CTS.TIMEDATE	TRT	MEAN.CIG.ALC.POT
1	54824 : 5	1:775	Min. :0.000	0:1900	Min. :1.00
2	54868 : 5	2:756	1st Qu.:0.469	1:1829	1st Qu.:1.00
3	54896 : 5	3:752	Median :1.499		Median :1.67
4	54919 : 5	4:731	Mean :1.590		Mean :1.85
5	54927 : 5	5:715	3rd Qu.:2.496		3rd Qu.:2.33
6	54945 : 5		Max. :3.811		Max. :5.00
7	(Other):3699				

Table 2. *Summary Table of the Clean Data*

Exploratory Analysis and Variograms

After cleaning and merging the data, 786 individuals were left in the data set. Note that not all individuals had all of the five measurements and there was only one obvious data recording error. Table 3 that there were 123 individuals with fewer than five observations. Most subjects had all five measurements, and only 10 individuals had one measurement out of five.

	1	2	3	4	5	6
Number of Observation	0	1	2	3	4	5
Number of Subjects	1	10	16	17	79	664

Table 3. *Number of subjects with a given observations for each subject*

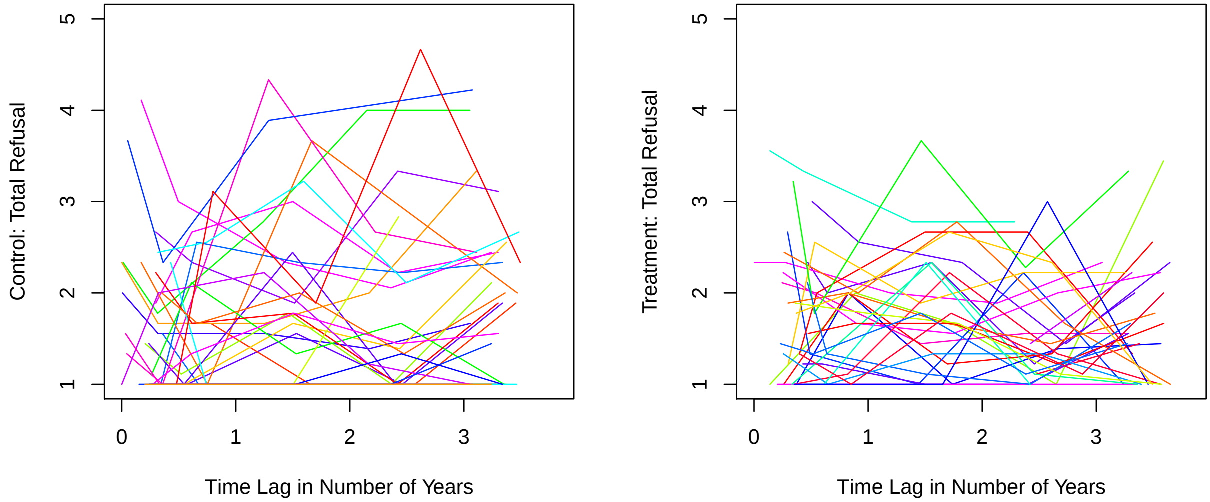
The following code chunk shows a basic way to print the total subject number, total observation number, and number of subjects per observation number (shown in Table 3) for data in long format.

```
cat(paste("Total number of subjects:", sum(table(unique(H145.long.comp$ID)))))
cat(paste("Total number of observations: ", length(H145.long.comp$ID)))
cat("Number of subjects with a given observations for each subject:")
table(table(H145.long.comp$ID))
```

After looking at the basic descriptives shown above, a good start for longitudinal EDA is to look at spaghetti plots for treatment and control groups separately. The following code chunk shows a simple way of creating spaghetti plots for treatment versus control groups.

```
## Spagetti plot of randomly chosen subjects for control group
## Randomly chosen seed for replicability purposes
set.seed(777)
plot( H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==0],
      H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==0],
      xlab="Time Lag in Number of Years",
```

```
     ylab="Control: Total Refusal", type="n")
## Sampleing 30 subjects
uid <- unique( H145.long.comp$ID[H145.long.comp$TRT==0] )
subset <- sample( uid, 30 )
for( j in 1:30 ){
  lines( H145.long.comp$CTS.TIMEDATE[ H145.long.comp$ID==subset[j] ],
         H145.long.comp$MEAN.CIG.ALC.POT[ H145.long.comp$ID==subset[j] ],
         col=sample(rainbow(30)) )
}
## Spagetti plot of randomly chosen subjects for treatment group
plot( H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==1],
      H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==1],
      xlab="Time Lag in Number of Years",
      ylab="Treatment: Total Refusal", type="n")
## Sampleing 30 subjects
uid <- unique( H145.long.comp$ID[H145.long.comp$TRT==1] )
subset <- sample( uid, 30 )
for( j in 1:30 ){
  lines( H145.long.comp$CTS.TIMEDATE[ H145.long.comp$ID==subset[j] ],
         H145.long.comp$MEAN.CIG.ALC.POT[ H145.long.comp$ID==subset[j] ],
         col=sample(rainbow(30)))
}
```



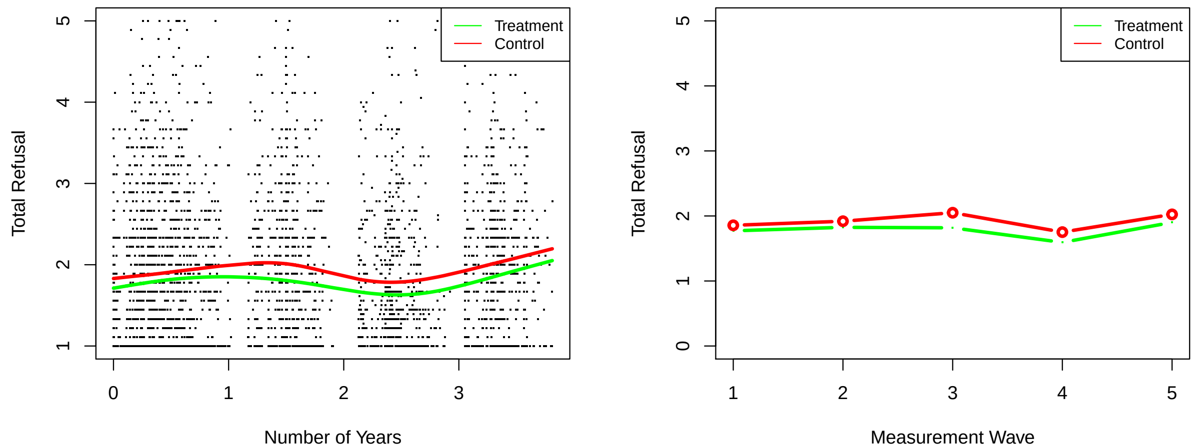
(a) *30 Randomly Choden Girls in Control Group* ***(b)*** *30 Randomly Choden Girls in Treatment Group*

Figure 1. *Spagetti plot of comparison between the treatment versus control group shows that the treatment group might have less variability and lower refusal score; which means the treatment group might be refusing more drug and alcohol*

Figure 1 shows that the variability of all the measurements for the control group was more pronounced compared to the treatment group. Also, in general, the refusal score was lower for the treatment group compared to the control; since total mean refusal skill was reverse coded, this means that the treatment group might be refusing drug and alcohol more often. To explore this phenomenon further, let us look at this observation more closely, first using a scatter plot along with smoothers, and second using the mean of each time lag for treatment and control groups (shown in the following code chunk).

```
# Higher df more detailed line, lower df closer to a line
plot( H145.long.comp$CTS.TIMEDATE, H145.long.comp$MEAN.CIG.ALC.POT,
      xlab="Number of Years",ylab="Total Refusal", pch=".")
```

```r
lines( smooth.spline( H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==1],
                      H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==1],
                      df=5 ), col="green", lwd=3)
lines( smooth.spline( H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==0],
                      H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==0],
                      df=5 ), col="red", lwd=3)
legend("topright",legend=c("Treatment", "Control"),
       col=c("green", "red"), lty=1, cex=0.8, bg="white")
## Calculating the mean of each lag
mc=c()
mt=c()
for(i in 1:5){
  mc[i]<- mean(H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==0&
                                                 H145.long.comp$time==i])
  mt[i]<- mean(H145.long.comp$MEAN.CIG.ALC.POT[H145.long.comp$TRT==1&
                                                 H145.long.comp$time==i])
}
## Plotting Means for Data Binned by Time Point
## for Treatment and Control Group separately
plot(mt, col="green", lwd=3,xlab="Measurement Wave",ylab="Total Refusal",
     type="b", ylim=c(0,5),pch=".")
points(mc, col="red", lwd=3, type = "b")
```



(a) *Smooth of cross-sectional total refusal over continous time for treatment versus control group* ***(b)*** *Means for data binned by time point for treatment and control group*

Figure 2. *Both of above plots shows that treatment group has consistently slightly lower refusal score (i.e. refusing more drug and alcohol)*

```
legend("topright",legend=c("Treatment", "Control"),

       col=c("green", "red"), lty=1, cex=0.8, bg="white")
```

To explore the systematic component of the total refusal score, one can examine the plots shown in Figure 2. Figure 2(a) is a scatter plot of total refusal over continuous time. The green smooth line is a spline with five degrees of freedom for the treatment group. The red smooth line is a spline with five degrees of freedom for the control group. These smooth splines fit a cubic smoothing spline between the knots to the total refusal score, forcing continuity at knots; this smooth spline can be thought of as a cross-sectional mean of total refusal score over time. Figure 2(b) shows the mean total refusal score at each time wave connected by lines. Both plots in Figure 2 show the same systematic pattern using two different methods. One can observe that the treatment and control groups start at pre-intervention with virtually the same refusal score; then the treatment group has a slight reduction all the way to the end. At the last measurement, the distance between the treatment and control groups decreases, which means the treatment effect might be disappearing.

Both cross-sectional plots in Figure 2 show the same rise and fall, so the researchers cannot be certain that the decrease in refusal score is due to the intervention. However, there could be a social or historical event affecting both groups around, for example, year 2 or 3 (i.e. the 4th measurement).

Another important part of the data to be explored is the number of observations at each measurement wave, as the accuracy of the estimations of different statistics at each time point depends on the number of observations. Sparse data issues can be investigated using Table 4; the R code corresponding to this table is as follows.

```
# Number of subjects per wave of measurements
obs.per.wave = matrix(NA, nrow = 1, ncol = 5)
colnames(obs.per.wave)<- c("Time1","Time2","Time3","Time4","Time5")
for(i in 1:5){
  obs.per.wave[1,i] = sum(table(unique(H145.long.comp$ID[H145.long.comp$time==i])))
}
```

	Time1	Time2	Time3	Time4	Time5
1	775	756	752	731	715

Table 4. *Number of subjects per wave of measurements*

Table 4 shows that we do not have sparse data issues since we have more than 700 observations at each time wave.

Introduction to Variogram Plots

The General Serial Covariance (GSC) model and its residual variabilities need to be understood in depth in order to plot and interpret a variogram. The GSC model was introduced in detail in the second paper, thus only a quick summary is provided here. The GSC model can be specified as follows:

$$Y_{ij} = \mu_{ij} + \alpha_i + W_i(t_{ij}) + \varepsilon_{ij} \tag{1}$$

where $i = 1, ..., N$ is the subject index and $j = 1, ..., n_i$ is the measurement index. Thus the GSC model has three sources of variation, namely, variation in the random intercept which comes from α_i, variation in the serial process which comes from $W_i(t_{ij})$, and variation in the measurement error which comes from ε_{ij}. In addition, it is often assumed that all these parameters are independent with:

$$var(\alpha_i) = \nu^2 \tag{2}$$

$$cov(W_i(t_{ij}), W_i(t_{ik})) = \tau^2 \rho(|t_{ij} - t_{ik}|) \tag{3}$$

$$var(\varepsilon_{ij}) = \sigma^2 \tag{4}$$

The serial correlation within the repeated measure is defined as an intrinsic stationary Gaussian process, $W_i(t_{ij})$, where,

- $E(W_i(t_{ij})) = 0$
- $cov(W_i(t_{ij}), W_i(t_{ik})) = \tau^2 \rho(|t_{ij} - t_{ik}|) = \tau^2 \rho(u)$ where u is the time lag between measurements for the same subject.

For example, one can specify $\rho(u) = e^{(-(\frac{u}{\phi})^c)}$ where $c = 1$ induces an Exponential serial correlation structure and $c = 2$ induces a Gaussian serial correlation structure. The rate of exponential decrease (sometimes called the range) is $\frac{1}{\phi}$. Another example would be a Linear serial correlation structure, which is defined as $\rho(u) = 1 - \frac{u}{d}$ for $u < d$ and zero otherwise. The range for the Linear serial correlation structure is defined as d, after which the correlation is assumed to be zero.

Note that all the spatial correlation formulas are presented using time lag u or time t and not in terms of space/location. The spatial correlation terminology was borrowed from geostatistics, where scholars deal with space/location. In order to use these spatial correlation functions, the concept of space is being converted to time for repeated measure data.

Many other functional forms can be defined for the serial correlation function. However, the Linear, Gaussian, and Exponential serial correlations are the most commonly used functions. The functional forms of the Exponential and Gaussian correlation structures are highly consistent with the data in Education and Psychology since in these disciplines, the correlation between the repeated measure decreases as the time lag increases. The Linear covariance structure, on the other hand, might not be as realistic for these disciplines since the correlation between the repeated measure will rarely go to zero abruptly. EDA for the covariance structure of the outcome utilizes these functional forms and the variabilities shown in Equations 2 to 4.

Before model fitting, it is essential to explore the covariance pattern using a variogram. Variograms offer an alternative function to autocorrelation function (ACF) plots that describe associations among repeated observations with *irregular observation* time. It is an exploratory tool that allows researchers to examine two aspects of the covariance structure, namely, functional form and the three variance components coming from the residual of the GSC model (shown in Equations 2 to 4). Given a stochastic process R, and time lag u, the variogram is defined as:

$$v(u) = \frac{1}{2}E[\{R(t) - R(t-u)\}^2], u \geq 0 \tag{5}$$

The function $v(u)$ is estimated by smoothing the scatter plot of the $\frac{1}{2}(R_{ij} - R_{ik})^2$ over $u_{jk} = |t_{ij} - t_{jk}|$. Note that for the GSC model shown in Equation 1, the residuals are defined as:

$$R_{ij}(u) = Y_{ij} - \mu_{ij} = \alpha_i + W_i(t_{ij}) + \varepsilon_{ij} \tag{6}$$

If $R(u)$ is stationary (i.e. the residual mean zero and equal variance of time points), the variogram is directly related to the autocorrelation function $\rho(u)$ via the following expression:

$$v(u) = \sigma^2 + \tau^2\{1 - \rho(u)\} \tag{7}$$

Equation 7 reveals:

- When the autocorrelation function $\rho(u)$ increases, the variogram $v(u)$ increases.
- As $u \to 0$ then $v(u) \to \sigma^2$.
- As $u \to \infty$ then $v(u) \to \sigma^2 + \tau^2$.

The total process variance for all individuals in the data can be written as:

$$\frac{1}{2}E[R_{ij} - R_{kl}] = \nu^2 + \tau^2 + \sigma^2, i \neq k \tag{8}$$

Thus, the total process variance shown in Equation 8 is estimated using the following expression:

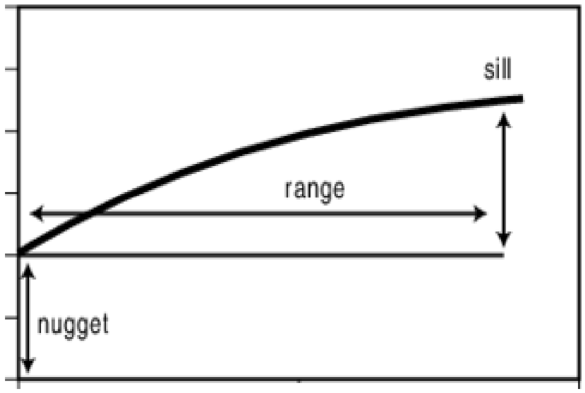
$$\hat{\nu}^2 + \hat{\tau}^2 + \hat{\sigma}^2 = \frac{1}{2N^*} \sum_{i \neq k} \sum_{i=1}^{n_i} \sum_{l=1}^{n_l} [R_{ij} - R_{kl}]^2 \tag{9}$$

where N^* is the number of terms in the sum. The estimate of total variance together with the variogram will be used for deciding which of the three stochastic components (shown in Equations 2 to 4) will be included in the model, and for selecting an appropriate serial correlation function $\rho(u)$. The former and latter are illustrated in Figures 3 and 4, respectively.

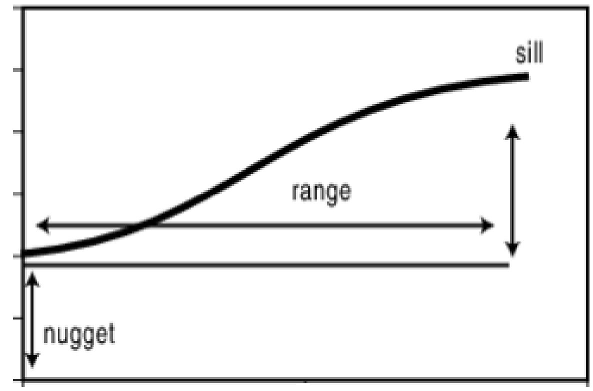
As shown in Figure 3, the shape of the variogram can help to identify the functional form of the covariance structure. A round increasing shape, an S-shape, and a linear shape are the most commonly occurring patterns that, respectively, indicate Exponential, Gaussian, and Linear covariance structures. Figure 4 illustrates that the height of the variogram can help with visualizing different components of the covariance structure. Looking at Figure 4 from bottom to top, one can label the variability in the measurement error, serial correlation, and random intercept as σ^2, τ^2, and ν^2, respectively. Note that the top horizontal dotted line in Figure 4 has been plotted using the total process variance estimated in Equation 9.

Note that there exist many complex spatial correlation structures, such as Rational Quadratic, Matern, and Spherical Correlation Structures, that are not presented in this paper. Readers who encounter variogram plots with shapes not covered in this paper can consult spatial data literature such as Carlin, Gelfand, and Banerjee (2014) to identify the structure of the repeated measure data shown in their variogram. Researchers can also combine and create new functions (Simpson, Edwards, Muller, Sen, & Styner, 2010) that suit their repeated measure correlation structure.

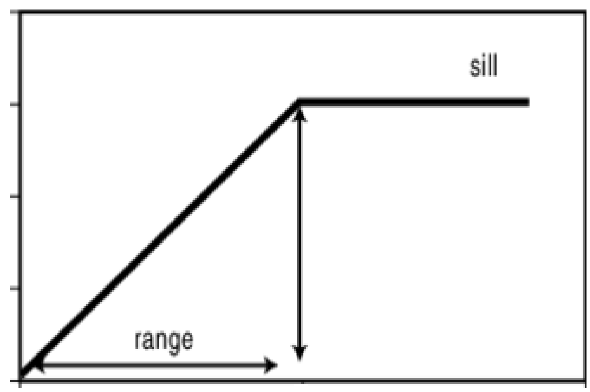
The next code chunk presents how to plot the variogram for the drug abuse prevention in-



***(a)** Exponential Variogram*



***(b)** Gaussian Variogram*



***(c)** Linear Variogram*

***Figure 3.** Variograms of different spatial correlation structures*

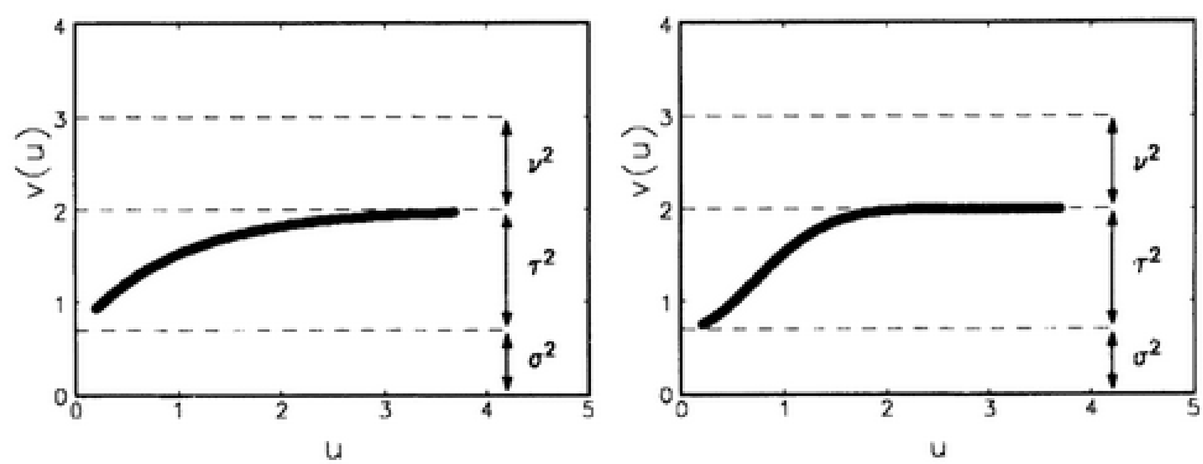


***Figure 4.** Sample variograms, showing* σ^2, τ^2, *and* ν^2, *which represent measurement error, serial correlation, and random intercept, respectively. Adopted from Verbeke and Molenberghs (2000, p. 143)*

tervention data. First, the outcome needs to be de-trended using a naive model such as a natural spline model (shown as `ns()` in the next code chunk). Second, the residuals are extracted from this linear spline model. Third, these residuals are used to explore the autocorrelation pattern. Note that the knots in the `ns()` function need to be chosen in terms of the time variable. One can place the knots at the 1st quantile, Median, and 3rd quantile of the CTS.TIMEDATE variable.

To compute an empirical variogram, the function `lda.variogram()` (written by P. Heagerty, shown in Appendix A) is used to plot the variogram shown in Figure 5. The shape of the variogram in Figure 5 can be judged as falling somewhere in between an Exponential and a Gaussian autocorrelation. If the middle dent in the variogram were deeper and S-shaped, this would be a variogram corresponding to the Gaussian serial correlation. On the other hand, if the middle dent were more rounded, this would be a variogram corresponding to the Exponential serial correlation.

The variability in the middle part of the plot (i.e. τ^2) shows that there is definitely some autocorrelation. The space between the top part of the variogram and the total variance dashed line (i.e. ν^2), shows that the model needs a random intercept. Finally, the gap between the bottom of the x-axis and the start of the plot (i.e. σ^2) shows that there is some leftover error (i.e. measurement error).

To sum up, in terms of model fitting, this variogram would indicate that the researcher should run two GSC models (both with random intercept and measurement error), one using a Gaussian serial correlation and one using Exponential. Then one can use ad-hoc criteria such as AIC, BIC, and Log-Likelihood to decide which serial correlation is optimal. However, before moving on to model fitting, one needs to check the variogram stationarity assumptions (i.e. residual mean zero and residual variances equal to each other).

```r
fit.3knots <- lm(MEAN.CIG.ALC.POT ~ ns(CTS.TIMEDATE,
    knots = c(quantile(H145.long.comp$CTS.TIMEDATE, na.rm = TRUE)[2],
        quantile(H145.long.comp$CTS.TIMEDATE, na.rm = TRUE)[3],
        quantile(H145.long.comp$CTS.TIMEDATE, na.rm = TRUE)[4])),
    data = H145.long.comp, na.action = NULL)
resids.3 <- residuals(fit.3knots)
H145.long.comp$resids.3 <- residuals(fit.3knots)
vario <- lda.variogram(id = H145.long.comp$ID, y = H145.long.comp$resids.3,
    x = H145.long.comp$CTS.TIMEDATE)
dr <- vario$delta.y
du <- vario$delta.x
tot.var.est <- var(H145.long.comp$resids.3)
plot(du, dr, pch = ".", ylim = c(0, 1.2 * tot.var.est),
    xlab = "Time Lag in Number of Years", ylab = "Variogram")
lines(smooth.spline(du, dr, df = 5), lwd = 3)
abline(h = tot.var.est, lty = 2, lwd = 2)
title("Total Refusal Residual Variogram")
```

The next code chunk illustrates one way to look into the covariance structure of the data, check the stationarity assumptions, and explore the sparse data issues; Table 5 to Table 7 show the output for this code chunk. Note that in order to use the commands in this code chunk, residuals need to be in *wide* format.

Table 5, which is the mean of residuals for each time point, shows that the means of residual time lags are all close to zero; this means the residual mean zero of the stationarity assumption is satisfied. The diagonal elements of the Table 6 show that the variance of the time lags are all roughly equal, thus the equal variance requirement of the stationarity assumption is satisfied as

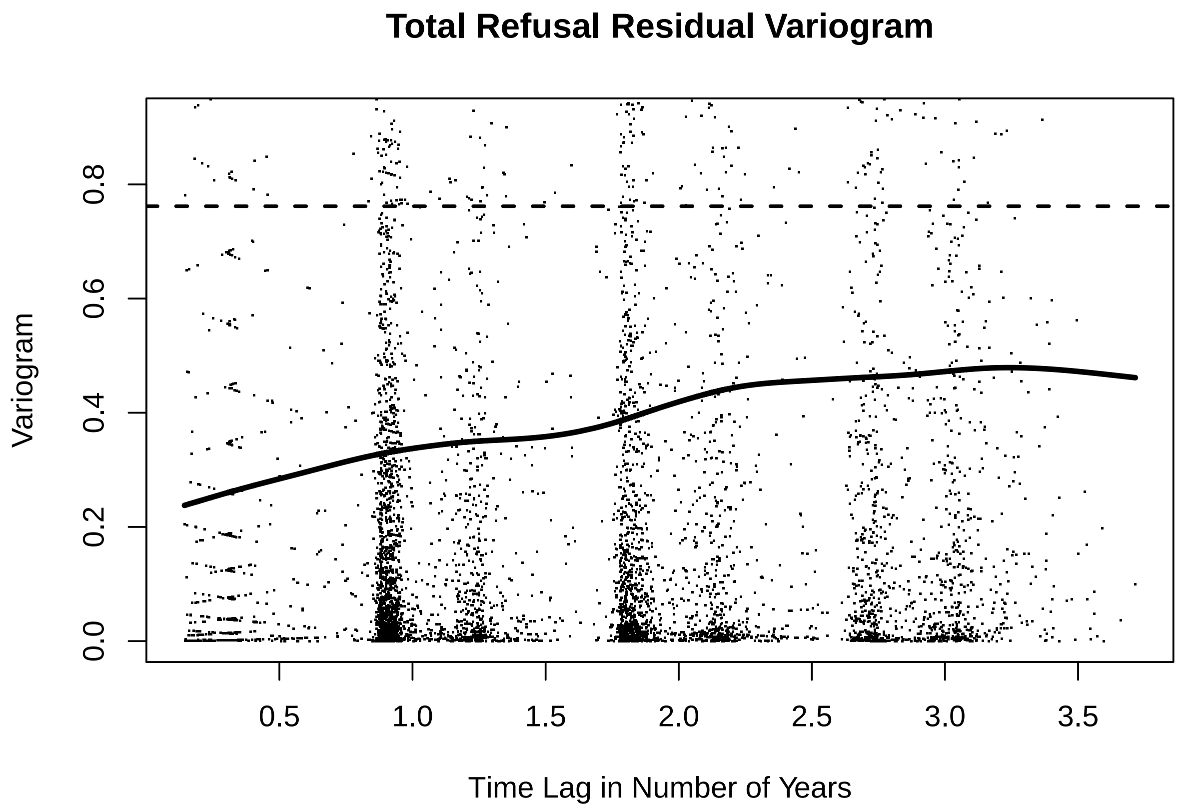


Figure 5. *Variogram of the residuals using a 3-knot natural spline. This looks like an Exponential or Gaussian correlation structure with measurement error, random intercept and auto-correlation. Note that the the y-limit has been manipulated for better resolution*

well. Together, Table 5 and Table 6 show that stationarity assumption is satisfied for the variogram. Note that Table 6 also indicates that as the time lag increases, the covariance between the repeated measure decreases, which is consistent with the Exponential/Gaussian covariance pattern shown in the variogram. Carlin et al. (2014) is a great in-depth source for researchers who are interested in studying variograms in more detail.

Finally, Table 7 shows the number of observations used to calculate each covariance shown in Table 6; the smallest number in this table is 685, which means sparse data issues are non-existent in this data set.

```r
## Residual mean equal 0?

colMeans(resids.wide, na.rm = TRUE)

## Residual variance equal to each other?

cov(resids.wide, use = "na.or.complete")

## Sparse data?

crossprod(!is.na(resids.wide))
```

	Resid.1	Resid.2	Resid.3	Resid.4	Resid.5
Mean	-0.0004	-0.0042	0.0147	-0.0272	0.0172

Table 5. *Mean of each lag is close to zero thus the mean zero part of the stationarity assupmtion is roughly satisfied*

	Resid.1	Resid.2	Resid.3	Resid.4	Resid.5
Resid.1	0.67	0.47	0.37	0.24	0.22
Resid.2	0.47	0.75	0.42	0.29	0.28
Resid.3	0.37	0.42	0.76	0.40	0.38
Resid.4	0.24	0.29	0.40	0.62	0.34
Resid.5	0.22	0.28	0.38	0.34	0.72

Table 6. *Covariance of residuals using a 3-knot natural spline. Looking at the diagonal elements, the variances are all roughly equal thus the equal variance part of the stationarity assumption is satisfired*

	Resid.1	Resid.2	Resid.3	Resid.4	Resid.5
Resid.1	775	756	752	720	708
Resid.2	756	756	743	711	698
Resid.3	752	743	752	712	696
Resid.4	720	711	712	731	685
Resid.5	708	698	696	685	715

Table 7. *Sparse data issues are not observed so the covarience parameters can be estimated well*

How to Run and Compare Different Models

In the previous section, it was established by using a variogram that a GSC model with either an Exponential or Gaussian covariance structure would be a good choice for the drug abuse prevention intervention data set. Table 8 shows the coefficients, standard errors (SE), and p-values from a random intercept model (i.e. HLM1), a random intercept and slope model (i.e. HLM2), a GSC model with Exponential covariance structure (i.e. Fit.Exp), and a GSC model with Gaussian covariance structure (i.e. Fit.Gauss). Detailed descriptions of HLM1 and HLM2 can be found in Singer and Willett (2003).

The next code chunk shows the R code corresponding to models shown in Table 8. The following are the model specifications used in this code chunk:

- `lme()` command can be used to run both the GSC and HLM models.
- The argument `method = "ML"` identifies that the log-likelihood is maximized. Alternatively, one can use restricted log-likelihood by utilizing the option `method = "REML"`.
- The argument `random = reStruct(~1 + CTS.TIMEDATE |ID, pdClass = "pdSymm")` is where random intercept and slope of time are defined. By eliminating the `+CTS.TIMEDATE` , one would be left with a random intercept-only model.
- The argument `data = H145.long.comp` defines which data set is being used. Note that in the `lme()` command, one needs to use the data in long format.
- The `control = lmeControl(opt = "optim")` option defines the estimation algorithm. Many different estimation algorithms are available in this option and users should choose the one that gives the least convergence issues.
- `summary()` command can be used to look at the details of the model fitting outcome.

- Readers who need more information about a specific method can type ? in the console along with a given command to learn more (e.g. ?lme()).

All the coefficient estimates of the models shown in Table 8 are known to be unbiased. The simulation study in Paper 2 indicated that HLM2 had lower power compared to the other three models, which can be inferred by referring to Table 8: looking at the p-values for HLM2, none of the variables other than intercept are significant. HLM1, on the other hand, has two significant variables out of three. However, one should not trust the SE from the HLM1 model since according to the simulation study, this model had a SE estimate furthest from the "true" SE (it was under-estimating).

Furthermore, the variogram identified the appropriateness of the GSC model with either Exponential or Gaussian covariance structures. AIC, BIC, and Log-Likelihood in Table 8 unanimously confirm that a GSC with Exponential covariance structure is an optimal choice; AIC and BIC are the smallest for this model, and Log-Likelihood is smallest in absolute value. Using these same criteria, the GSC model with a Gaussian covariance structure is the second best model; note that for these two models that we trust the most (since according to the simulation study they have the correct SE and are more powerful), we only have two significant coefficients.

```
## Running different models The variogram looked like
## a exponential covariance structure Thus the
## corEXP() should be performin the best

## Random intercept only model called HLM1
HLM1 = lme(MEAN.CIG.ALC.POT ~ CTS.TIMEDATE + TRT + CTS.TIMEDATE *
    TRT, method = "ML", random = reStruct(~1 | ID, pdClass = "pdSymm"),
    data = H145.long.comp, control = lmeControl(opt = "optim"))
```

```r
summary(HLM1)

## Random intercept and slope model called HLM2

HLM2 = lme(MEAN.CIG.ALC.POT ~ CTS.TIMEDATE + TRT + CTS.TIMEDATE *

    TRT, method = "ML", random = reStruct(~1 + CTS.TIMEDATE |

    ID, pdClass = "pdSymm"), data = H145.long.comp, control = lmeControl(opt = "optim"))

summary(HLM2)

## Random intercept model plus Linear correlation

## structure

fit.lin = lme(MEAN.CIG.ALC.POT ~ CTS.TIMEDATE + TRT +

    CTS.TIMEDATE * TRT, method = "ML", random = reStruct(~1 |

    ID, pdClass = "pdSymm"), correlation = corLin(form = ~CTS.TIMEDATE |

    ID, nugget = TRUE), data = H145.long.comp, control = lmeControl(opt = "optim"))

summary(fit.lin)

## Random intercept model plus Exponential correlation

## structure

fit.exp = lme(MEAN.CIG.ALC.POT ~ CTS.TIMEDATE + TRT +

    CTS.TIMEDATE * TRT, method = "ML", random = reStruct(~1 |

    ID, pdClass = "pdSymm"), correlation = corExp(form = ~CTS.TIMEDATE |

    ID, nugget = TRUE), data = H145.long.comp, control = lmeControl(opt = "optim"))

summary(fit.exp)

## Random intercept model plus Gaussian correlation

## structure

fit.gaus = lme(MEAN.CIG.ALC.POT ~ CTS.TIMEDATE + TRT +

    CTS.TIMEDATE * TRT, method = "ML", random = reStruct(~1 |

    ID, pdClass = "pdSymm"), correlation = corGaus(form = ~CTS.TIMEDATE |

    ID, nugget = TRUE), data = H145.long.comp, control = lmeControl(opt = "optim"))

summary(fit.gaus)
```

Appendix A shows functions that can be used to extract parameters from the `lme()` function. If the variance parameters are of interest, one can use the equivalency formulas derived in Appendix C to extract the parameters manually. These formulas have been implemented in the function shown in Appendix A called `extract_HLM` and `extract_lme`. The next code chunk shows how to extract the estimated coefficient and variance parameters from the above output using the functions shown in Appendix A; corresponding formulations are derived in Appendix C. For more details on how to extract parameters from the R output, see Martinussen, Skovgaard, and Sorensen (2012).

```
extract_HLM(HLM1)

extract_HLM(HLM2)

extract_lme(fit.exp)

extract_lme(fit.gaus)
```

The GSC model was introduced in the Introduction to Variogram Plots section. The fixed-effect part (i.e. μ_{ij}) of the GSC model for this analysis can be written as:

$$\begin{aligned} Y_{ij} &= \mu_{ij} + \alpha_i + W_i(t_{ij}) + \varepsilon_{ij} \\ \mu_{ij} &= \beta_0 + \beta_1 Time + \beta_2 Treatment + \beta_3 Time \times Treatment \end{aligned} \tag{10}$$

The coefficients of the model shown above are estimated with different models in Table 8. Using the estimations from the GSC model with an Exponential covariance structure, the fixed-effect coefficients can be interpreted as follows:

- Since there is interaction in the model, readers need to be careful when interpreting the main effects as the interaction might be driving the main effect.
- The estimated intercept $\beta_0 = 1.8809$ refers to the population total refusal score when $Time =$

		HLM1	HLM2	Fit.Exp	Fit.Gauss
(Intercept)	Estimate	1.8917^{***}	1.8932^{***}	1.8809^{***}	1.8757^{***}
	SE	0.0399	0.0448	0.0431	0.0425
	P-Value	0.0000	0.0000	0.0000	0.0000
CTS.TIMEDATE	Estimate	0.0278^{*}	0.0255	0.0342^{*}	0.0374^{*}
	SE	0.0128	0.0157	0.0157	0.0150
	P-Value	0.0291	0.1044	0.0289	0.0127
TRT1	Estimate	-0.1150^{*}	-0.1147	-0.1130	-0.1161
	SE	0.0567	0.0637	0.0611	0.0603
	P-Value	0.0430	0.0721	0.0649	0.0544
CTS.TIMEDATE:TRT1	Estimate	-0.0162	-0.0160	-0.0144	-0.0148
	SE	0.0181	0.0223	0.0222	0.0212
	P-Value	0.3708	0.4729	0.5152	0.4866
AIC		8458.9342	8341.3526	8320.7047	8334.5976
BIC		8496.2776	8391.1437	8370.4958	8384.3888
Log Likelihood		-4223.4671	-4162.6763	-4152.3523	-4159.2988
Num. obs.		3729	3729	3729	3729
Num. groups		786	786	786	786

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8. *Comparison of the statistical models with respect to the estimates, standard error (SE) and P-values. Note that HLM1 denotes the random intercept only model and HLM2 denotes the random intercept and slope model. Fit.Exp and Fit.Gauss denote the GSC model with Exponential and Gaussian covariance structures, respectively*

0 and $Treatment = 0$. In other words, in the population, at time zero for the control group, the estimated refusal score is 1.8809.

- The estimated parameter $\beta_1 = 0.0342$ refers to the population yearly rate of change in total refusal score for the control group. For the treatment group (i.e. $Treatment = 1$), the effect of Time is $\beta_1 \times Time + \beta_3 \times Time = (\beta_1 + \beta_3) \times Time$.
- The estimated parameter $\beta_2 = -0.1130$ refers to the population average difference in total refusal score at time zero, comparing the treatment group to the control group. For all other

times, the population average difference in total refusal score, comparing the treatment group to the control group, is $\beta_2 + \beta_3 \times Time$.

- The estimated parameter $\beta_3 = -0.0144$ can be interpreted in two different ways: (a) the effect of treatment on the total refusal score is estimated to be varying over time by -0.0144 in the population; and (b) the population rate of change of total refusal score over time comparing the treatment to the control group has been estimated to be -0.0144 (this coefficient is the difference in the slope parameter as time varies between the treatment and control groups).

Note that each parameter's interpretation is using a *conditional* expectation; each parameter's interpretation is contingent on keeping other variables constant. For more details on interpretation of the multilevel modeling coefficients see Chapter 5 of Singer and Willett (2003).

Table 8 can also be used to compare the estimated coefficients, SEs, and p-values of each variable across the different models. For example, comparing the estimations corresponding to CTS.TIMEDATE for HLM2 and Fit.Exp, one can observe the following:

- Regarding the parameter estimation, we observe an approximate 34% (from 0.0255 to 0.0342) increase when comparing HLM2 to Fit.Exp. However, the corresponding confidence intervals do overlap (i.e. HLM2=[-0.0059, 0.0569] versus Fit.Exp=[0.0028, 0.0656]).
- From the simulation study, we know that the SE for both models is about 0.0157–close to the "true" SE.
- In the simulation study, we have observed that HLM2 has lower power compared to the GSC models. Here, the p-value corresponding to HLM2 is non-significant and the p-value corresponding to the Fit.Exp model is significant.

By making similar observations for all the coefficients and estimations, one can conclude that the choice of modeling can affect the estimation and testing of the fixed-effect parameters.

Utilizing the variogram and the all of the model fit criteria, it is now established that the GSC model with Exponential covariance structure is the optimal model. Figure 6 corresponds to the next code chunk, which is the plot of the fixed-effect linear trend corresponding to the GSC model with an Exponential covariance structure. Researchers can refer to Figure 6 to explore the linear trend visually. The total refusal score for the control group is consistently higher, compared to the treatment group, with a slightly increasing pattern overall. This linear trend would be relatively restrictive if researchers believe that the effect of treatment might be different in various time intervals. For example, the effect of treatment on refusal score might be steeper in the first year compared to the last year, but this model estimates one slope for the entire time. As shown in the next section, using different polynomial regression curves for different time intervals can model the fixed effect in a more flexible way, but this method renders the parameter estimates uninterpretable.

```
# Plot of fitted values from the GSC model with Exponential covariance structure
plot(H145.long.comp$CTS.TIMEDATE, H145.long.comp$MEAN.CIG.ALC.POT,
     xlab="Time in Number of Years", ylab="Total Refusal Score",pch=".")
fitted.mean.GSC.Exp <- fit.exp$fitted[, 1]
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==0]),
       sort(fitted.mean.GSC.Exp[H145.long.comp$TRT==0]), col="red", lwd=2 )
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==1]),
       sort(fitted.mean.GSC.Exp[H145.long.comp$TRT==1]), col="green", lwd=2 )
legend("topright",legend=c("Control", "Treatment"),
       col=c("red","green"), lty=1, cex=0.7, bg="white")
```

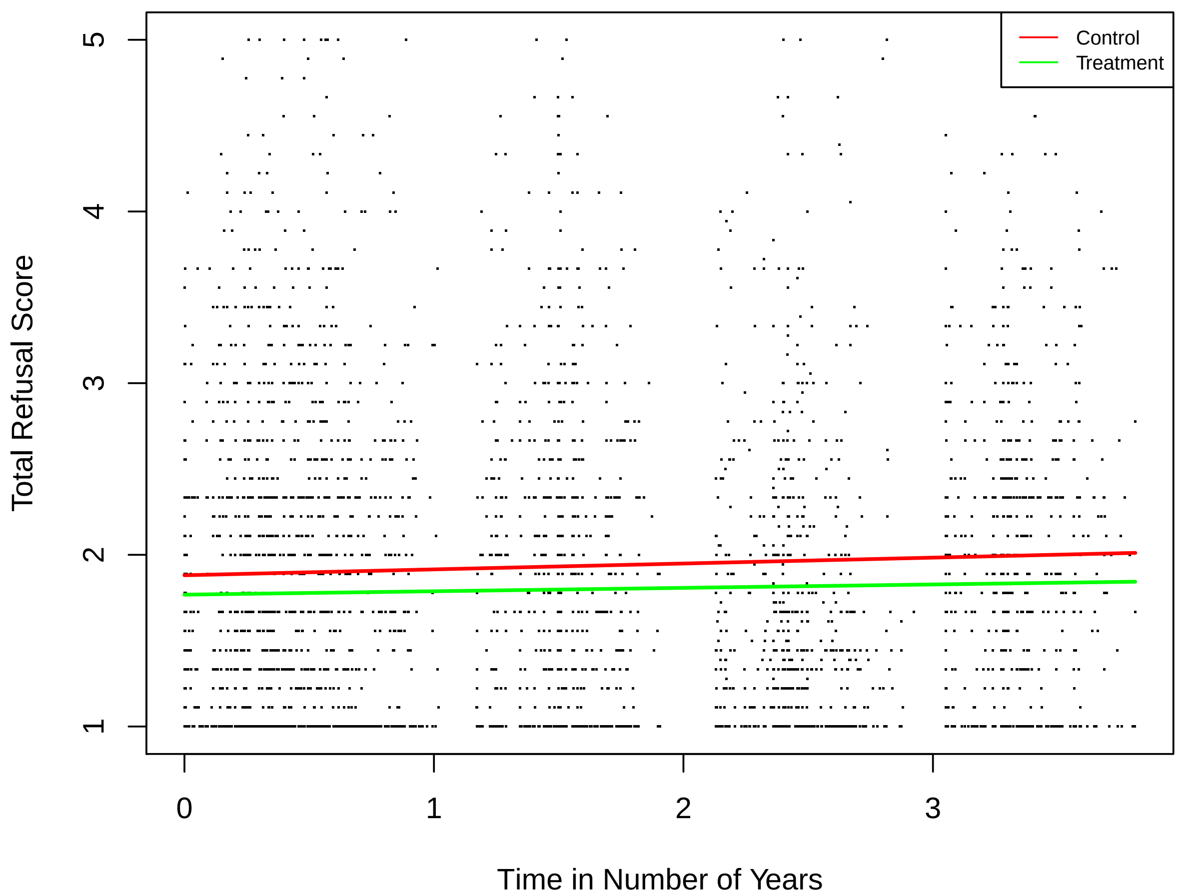


Figure 6. *Plot of the fitted values of the GSC model with Exponential covariance structure for treatment and control group separately*

Modeling the Fixed Effect More Flexibly

So far, a multiple linear regression formulation was used to model the fixed effect of the GSC model. The fixed-effect part can be modeled more flexibly using spline models; these approaches use a set of basis functions to fit a parametric model. Although there are many different types of splines, in this section Basis Splines (i.e. B-splines) and Natural Splines (N-Splines) will be covered. A brief discussion of splines is provided in this section. Readers interested in more details on these types of modeling techniques can refer to books such as Chambers, Hastie, et al. (1992); Friedman, Hastie, and Tibshirani (2001); Maindonald and Braun (2006); Yee (2015); and Dunn and Smyth (2018).

A spline curve can be thought of as joining different piece-wise polynomials together at different thresholds called *knots*. Knots can be defined systematically or intuitively; one can choose the knots to be at different quantiles or decide where to put the knots based on the data. The piece-wise polynomial that runs between the knots can be of any degree, however, the focus of this section will be on degree three polynomials. This is because degree three polynomials are not too complex for Education and Psychology data, yet are flexible and widely used. To prevent rigidness at the knots (where the splines are connected), B-splines and N-splines constrain the first and second derivatives. While B-splines do not implement constraints at the boundaries (i.e. min and max of the data), N-splines implement constraints such that linearity is assured beyond boundaries. The N-splines are usually a better choice in terms of their behavior at boundaries and, due to the boundary constraints, N-splines have two extra degrees of freedom compared to B-splines.

There exist different formulations for presenting B-splines, N-splines, and the implemented constraints. Note that an N-spline is just a B-spline with extra boundary conditions. Using B-splines or N-splines of the time variable in the GSC Exponential model is the same as transforming the time variable into a set of B-splines. Using this basis matrix instead of the original variable adds flexibility to the modeling of the fixed effect. Given k interior knots and d, the degree of the piece-wise polynomial, the general form of the basis function can be written as follows:

$$f(x) = \beta_0 + \sum_{j=1}^{k+d} \beta_j B_j(x) \tag{11}$$

The focus of this section is on cubic splines (i.e. $d = 3$).

The next code chunk shows the details of running a GSC model with Exponential covariance structure and splines of the time variable in the fixed part. The biggest trade-off in using splines is that the coefficients of this model are not interpretable, whereas those of a regular multiple linear regression are. The plot of the fixed effect is mainly used to visually analyze the pattern of the data.

The next code chunk shows one way of adding spline basis to the GSC model with an Exponential covariance structure. The R code description is similar to the previous section. The only difference is that now, instead of having the time variable, one uses the B-spline (i.e. `bs()`) or N-spline (i.e. `ns()`) of the time variable with knots defined at the 25%, 50%, and 75% quantiles. The argument `intercept = FALSE` indicates that the defined basis does not need to include the intercept since the `lme()` command already includes it in the model. The degree of B-spline basis is simply the degree of the piece-wise polynomial to be fit, which is defined as a degree three polynomial here. However, one can use a higher or lower degree polynomial if needed. For the N-spline, this degree is automatically set to three so it is not defined in the R code.

```
## Random intercept model plus Exponential correlation structure
## plus B-Spline and N-Spline for CTS.TIMEDATE
## Implementing four knots at using the quantiles of CTS.TIMEDATE
fit.exp.bs.4knots = lme( MEAN.CIG.ALC.POT ~ bs(CTS.TIMEDATE,
                                               knots=c(0.469, 1.590, 2.496, 3.411),
                                               degree=3, intercept = FALSE ) + TRT +
                         bs(CTS.TIMEDATE, knots=c(0.469, 1.590, 2.496, 3.411),
                            degree=3, intercept = FALSE)*TRT,
                        method = "ML", random = reStruct( ~ 1 | ID, pdClass="pdSymm"),
```

```
                         correlation = corExp( form = ~ CTS.TIMEDATE | ID, nugget=TRUE),
                         data =  H145.long.comp, control=lmeControl(opt="optim"))
summary(fit.exp.bs.4knots)
fit.exp.ns.4knots =  lme( MEAN.CIG.ALC.POT ~ ns(CTS.TIMEDATE,
                                              knots=c(0.469, 1.590, 2.496, 3.411),
                                              intercept = FALSE )
                          + TRT + ns(CTS.TIMEDATE,
                                   knots=c(0.469, 1.590, 2.496, 3.411),
                                   intercept = FALSE)*TRT,
                          method = "ML", random = reStruct( ~ 1 | ID, pdClass="pdSymm"),
                          correlation = corExp( form = ~ CTS.TIMEDATE | ID, nugget=TRUE),
                          data =  H145.long.comp, control=lmeControl(opt="optim"))
summary(fit.exp.ns.4knots)
```

The next code chunk shows the details of how to plot the B-spline and the N-spline for comparison and analysis purposes. Figure 7 is the resulting plot from this code chunk. The blue vertical lines are the knots chosen at the 25%, 50%, and 75% quantiles. Figure D1 in Appendix D shows the same model but with a different set of knots. In Figure D1 the knots are chosen based on the different waves of data collection (i.e. at 0.268493, 1.268493, 2.268493, 3.268493 years). The choice of knots did not change the results; the former and latter plots are almost identical and both show a constant increase in total refusal score between the knots and overall. However, the initial increase between knots is less steep and the slope gets steeper as we get to the final waves of data. Researchers can use the knots to identify at which points an intervention booster might be appropriate.

```r
fitted.mean.bs.4knots <- fit.exp.bs.4knots$fitted[,1]
plot(H145.long.comp$CTS.TIMEDATE, H145.long.comp$MEAN.CIG.ALC.POT,
     xlab="Time in Number of Years", ylab="Total Refusal Score", pch=".")
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==0]),
       sort(fitted.mean.bs.4knots[H145.long.comp$TRT==0]), col="red", lwd=2, lty=2 )
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==1]),
       sort(fitted.mean.bs.4knots[H145.long.comp$TRT==1]), col="red", lwd=2 )
fitted.mean.ns.4knots <- fit.exp.ns.4knots$fitted[,1]
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==0]),
       sort(fitted.mean.ns.4knots[H145.long.comp$TRT==0]), col="green", lwd=2, lty=2 )
lines( sort(H145.long.comp$CTS.TIMEDATE[H145.long.comp$TRT==1]),
       sort(fitted.mean.ns.4knots[H145.long.comp$TRT==1]), col="green", lwd=2 )
abline(v=c(0.469, 1.590, 2.496, 3.411), col="blue")
legend("topright",legend=c("Control.bs", "Treatment.bs","Control.ns", "Treatment.ns"),
       col=c("red", "red","green","green"), lty=c(2,1,2,1), cex=0.7,bg="white")
```

Note that all the information needed for model comparison is given in the `summary()` output. However, for a concise model comparison along with testing nested models, one can use the `anova()` command as shown in the following code chunk.

```r
# Model comparison
# Note Because you want to maximize the log-likelihood, the higher value is better.
# For example, a log-likelihood value of -3 is better than -7.
anova(HLM1, HLM2, fit.lin, fit.exp, fit.gaus, fit.exp.bs.4knots, fit.exp.ns.4knots)
```

Finally Table 9 is a quick comparison of B-splines versus N-splines. A fun fact from Yee (2015) about splines is a great way to end this section:

The word 'spline' comes from a thin flexible strip used by engineers and architects in the

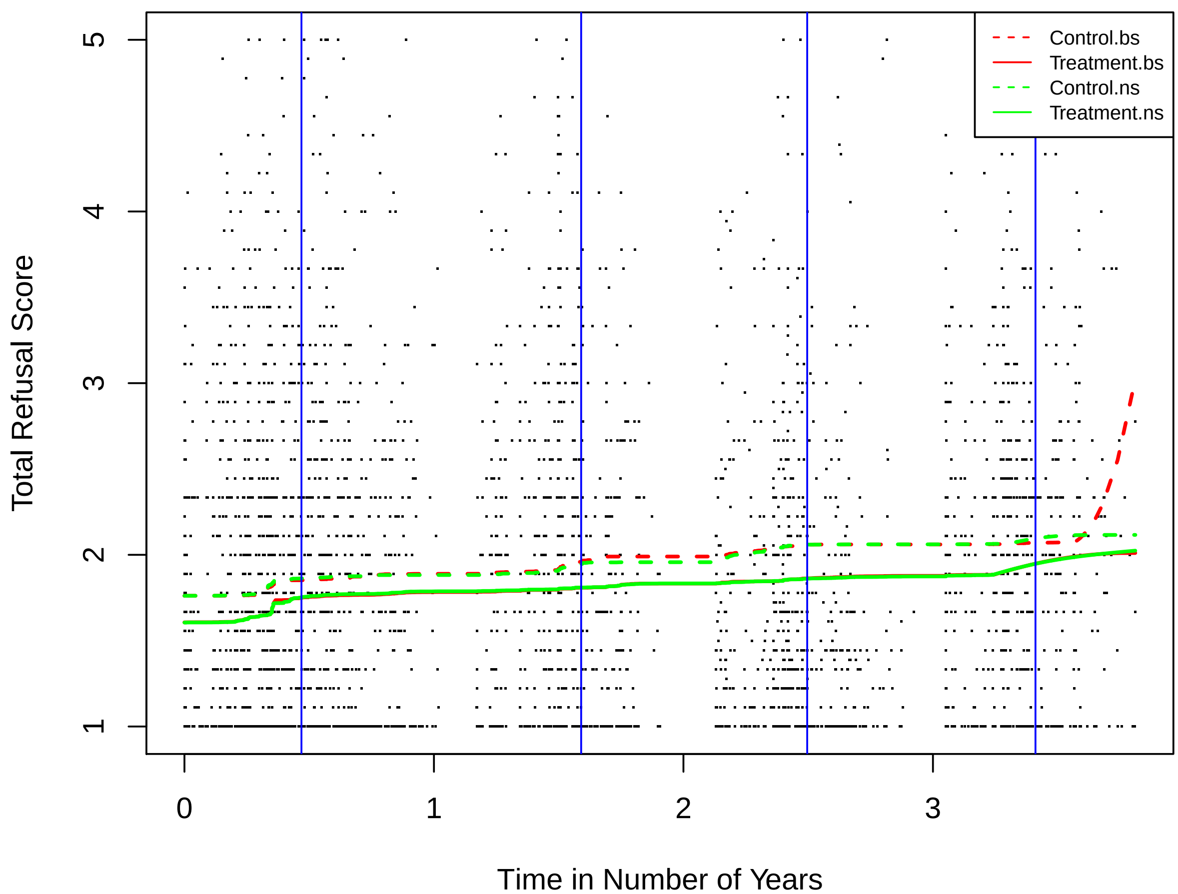


Figure 7. *Plot of the fitted values of the B-spline and N-spline for treatment and control group separately. Knots are placed at second, third and fourth quantiles.*

pre-computer days to construct ship hulls and the hydrofoils of wings. Splines were attached to important positions on a 2-dimensional plan (e.g., floor of a design loft or on enlarged graph paper) using lead weights called "ducks"and then released. The resting shapes assumed by the splines would minimize the strain energy according to some calculus of variations criterion. Splines were used by ancient Greek mathematicians (including Diocles) for drawing curves in diagrams (e.g., conic sections). In more modern times, I. J. Schoenberg is attributed to be the first to use 'splines' in the mathematical literature, and is known as the father of splines. The physical meaning of splines is especially relevant to the smoothing spline, where it is related to curvature and Hooke's Law for elastic bodies such as springs.

	B-spline	N-Spline
Definition	Generates a B-spline basis for a polynomial spline	Generates a B-spline basis for a natural cubic spline
Degree	Any degree of piece-wise polynomial	Only degree 3 piece-wise polynomial
Degrees of Freedom	Number of regression coefficients used to fit a regression spline = number of internal knots +3	Number of regression coefficients used to fit a regression spline = number of internal knots +1
Constraints at Knots	Smoothness at knots is assured by constraining the $d-1$ derivatives to be continuous at the knots when d is the degree of the piece-wise polynomial	Smoothness at knots is assured by constraining the first two derivatives to be continuous at the knots
Boundary Constrains	No constraints at boundaries thus unpredictable behavior at the end points	Second derivative forced to zero at end points thus linear at the end points

Table 9. *Comparison of the B-spline versus N-spline*

Discussion

It is important to use HLM approaches in order to explicitly take into account the structure of the data. In addition to accounting for the multilevel of the data when repeated measure data are used, the covariance structure of the repeated measure needs to be incorporated thoughtfully into modeling in order to perform proper analysis. The EDAs (implemented in R) presented in this tutorial can help the reader to explore the covariance structure of the data at hand before diving into modeling. This paper has suggested that researchers utilize the GSC model with a spatial covariance structure, and consult a variogram to identify the functional form and variability components of the covariance structure. Finally, model fitting and ad-hoc model comparison have been presented with an additional step to model the fixed effect more flexibly using splines.

Undoubtedly, the GSC and HLM models have more to offer than this tutorial has covered. However, this paper is a good starting point for researchers interested in using the GSC model in R. Note that the focus of this paper is on EDA (and the use of variograms), model fitting, and ad-hoc model comparison. However, assumption checking and model diagnostics are essential steps that are reserved for a future paper.

References

Carlin, B. P., Gelfand, A. E., & Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL: Chapman and Hall/CRC.

Chambers, J. M., Hastie, T. J., et al. (1992). *Statistical models in S* (Vol. 251). Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Dunn, P. K., & Smyth, G. K. (2018). *Generalized linear models with examples in R*. New York, NY: Springer.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1) (No. 10). New York, NY: Springer.

Maindonald, J., & Braun, J. (2006). *Data analysis and graphics using R: an example-based approach* (Vol. 10). New York, NY: Cambridge University Press.

Martinussen, T., Skovgaard, I. M., & Sorensen, H. (2012). *A first guide to statistical computations in R*. Samfundslitteratur.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/.

Schwinn, T., Hopkins, J., Schinke, S. P., & Liu, X. (2017). Using Facebook ads with traditional paper mailings to recruit adolescent girls for a clinical trial. *Addictive Behaviors*, *65*, 207–213.

Simpson, S. L., Edwards, L. J., Muller, K. E., Sen, P. K., & Styner, M. A. (2010). A linear exponent AR(1) family of correlation structures. *Statistics in Medicine*, *29*(17), 1825–1838.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New Yoek, NY: Oxford University Press.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York, NY: Springer Science & Business Media.

Yee, T. W. (2015). *Vector generalized linear and additive models: with an implementation in R*. New York, NY: Springer.

Appendices

Appendix A
Preliminary Steps Before Running Analysis

```
# Clearing the R memory
rm(list=ls())

# Options to be used through out the analysis
options(width=60)
options(scipen=1, digits=4)

# Setting working directory
setwd("C:/Users/SONY/Desktop/Paper3-V2")

# What's in the directory?
dir()

# Libraries used to create this document
library(knitr)
library(xtable)
library(pander)
library(MASS)
library(nlme)
library(splines)
library(joineR)
library(doBy)
library(stats)
library(dplyr)
library(graphics)
library(formatR)
library(rdd)
library(texreg)

# Loading the data
load("C:/Users/SONY/Desktop/Paper3-V2/H1.Rdata")
load("C:/Users/SONY/Desktop/Paper3-V2/H4.Rdata")
load("C:/Users/SONY/Desktop/Paper3-V2/H5.Rdata")

# What just have been loaded?
ls()

#-------------------------------------------------------------------
# Function to compute empirical variogram for continuous longitudinal data
# Author: Dr.Patrick Heagerty, retrieved from Dr. Daniel Gillen's lecture notes
# INPUT:  id = (nobs x 1) id vector
#          y = (nobs x 1) response (residual) vector
#          x = (nobs x 1) covariate (time) vector
#
# RETURN:  delta.y = vec( 0.5*(y_ij - y_ik)^2 )
#          delta.x = vec( abs( x_ij - x_ik ) )
```

```r
lda.variogram <- function( id, y, x ){
  uid <- unique( id )
  m <- length( uid )
  delta.y <- NULL
  delta.x <- NULL
  did <- NULL
  for( i in 1:m ){
    yi <- y[ id==uid[i] ]
    xi <- x[ id==uid[i] ]
    n <- length(yi)
    expand.j <- rep( c(1:n), n )
    expand.k <- rep( c(1:n), rep(n,n) )
    keep <- expand.j > expand.k
    if( sum(keep)>0 ){
      expand.j <- expand.j[keep]
      expand.k <- expand.k[keep]
      delta.yi <- 0.5*( yi[expand.j] - yi[expand.k] )^2
      delta.xi <- abs( xi[expand.j] - xi[expand.k] )
      didi <- rep( uid[i], length(delta.yi) )
      delta.y <- c( delta.y, delta.yi )
      delta.x <- c( delta.x, delta.xi )
      did <- c( did, didi )
    }
  }
  out <- list( id = did, delta.y = delta.y, delta.x = delta.x )
  out
}

# Function for extracting parameters for GSC ---------------------------
extract_lme = function(fit) {
  nugget = coef(fit$modelStruct$corStruct, unconstrained = F)[2]
  residual = fit$sigma

  sigma_hat = sqrt(residual^2*nugget)
  tau_hat = sqrt(residual^2 - sigma_hat^2)
  phi_hat = coef(fit$modelStruct$corStruct, unconstrained = F)[1]
  nu_hat = as.numeric(VarCorr(fit)[1, 2])


  value =  c(as.numeric(summary(fit)$tTable[, 1]),
             as.numeric(summary(fit)$tTable[, 2]),
             nu_hat, sigma_hat, tau_hat, phi_hat)


  names(value) = c(rownames(summary(fit)$tTable),
                   paste0(rownames(summary(fit)$tTable), "_se"),
                   "nu", "sigma", "tau", "phi")
  return(value)
}

# Function for extracting parameters for HLM -------------------------
extract_HLM = function(fit) {
```

```
  value =  c(as.numeric(summary(fit)$tTable[, 1]),
             as.numeric(summary(fit)$tTable[, 2]))

  names(value) = c(rownames(summary(fit)$tTable),
                   paste0(rownames(summary(fit)$tTable), "_se"))
  return(value)
}
```

Appendix B
Refusal Skill Questionnaire

The outcome variable (i.e. total refusal skill) has been created taking the average of the TotALCREF, TotCIGREF, and TotPOTREF, using the questionnaires shown in Table B1.

ALCOHOL REFUSAL SKILLS	
AlcRef1	Tell them "no" or "no thanks?"
AlcRef2	Tell them not now?
AlcRef3	Change the subject?
AlcRef4	Tell them you don't want to do it?
AlcRef4	Make up an excuse and leave?
TotALCREF	Mean of AlcRef1, AlcRef4, and AlcRef5
CIGARETTE REFUSAL SKILLS	
CigRef1	Tell them "no" or "no thanks?"
CigRef2	Tell them not now?
CigRef3	Change the subject?
CigRef4	Tell them you don't want to do it?
CigRef5	Make up an excuse and leave?
TotCIGREF	Mean of CigRef1, CigRef4, and CigRef5
MARIJUANA REFUSAL SKILLS	
PotRef1	Tell them "no" or "no thanks?"
PotRef2	Tell them not now?
PotRef3	Change the subject?
PotRef4	Tell them you don't want to do it?
PotRef5	Make up an excuse and leave?
TotPOTREF	Mean of PotRef1, PotRef4, and PotRef5

Table B1. *Original questionnaires for data collection regarding cigarette, alcohol, and marijuana refusal skills.*

The answers to all of the questions shown in Table B1 were inverse coded as follows:

- -99999 = Prefer not to answer.
- 1 = Definitely would.
- 2 = Most likely would.
- 3 = Not sure.
- 4 = Most likely would not.
- 5 = Definitely would not.

Appendix C

Equivalency Formulas for Variance Components

The random effects section of the output has the following two components:

- StdDev of the Intercept which is the estimated square root of the variance associated with the random intercept (i.e. ν).
- StdDev of the Residual which is the estimated square root of $\tau^2 + \sigma^2$.

Parameter estimate(s) section of the output has the following two components:

- Range which is the estimated parameter ϕ in the functional form of the serial correlation.
- Nugget which is the estimated quantity for $\frac{\sigma^2}{\tau^2+\sigma^2}$.

The estimated Nugget and StdDev of the Residual together are used to derive the variance associated with the measurement error as follows:

$$nugget = \frac{\sigma^2}{\tau^2 + \sigma^2} \Longrightarrow \sigma^2 = nugget \times (\tau^2 + \sigma^2) \tag{12}$$

Finally, utilizing σ^2, which was derived in Equation 12, the variance associated with the serial correlation (i.e. τ^2) can be extracted by substituting σ^2 in the Residuals StdDev (which was given in the output) as follows:

$$ResidualsStdDev = \sqrt{\tau^2 + \sigma^2} \Longrightarrow \tau^2 = (ResidualsStdDev)^2 - \sigma^2 \tag{13}$$

Appendix D
Knot Selection Based on Data Collection Time

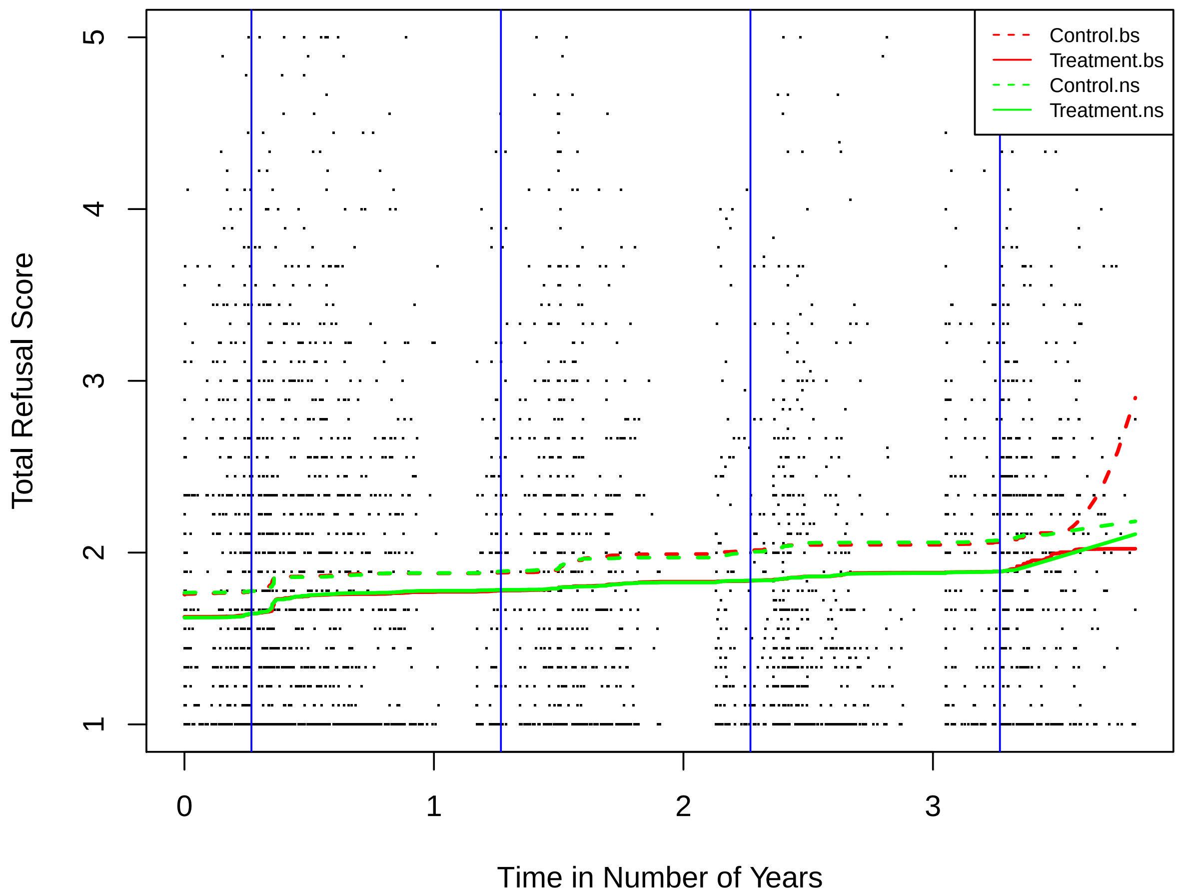


Figure D1. *Plot of the fitted values of the B-spline and N-spline for treatment and control group separately. Knots are placed based on data collection waves*