2014

A Monte Carlo Study: The Consequences of the Misspecification of the Level-1 Error Structure

Merlande Petit-Bois
University of South Florida, petitbois.m@gmail.com

A Monte Carlo Study: The Consequences of the Misspecification of the Level-1 Error Structure when Meta-Analyzing Single-Case Designs

by

Merlande Petit-Bois

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
in Curriculum and Instruction
with an emphasis in Measurement and Evaluation
Department of Educational Measurement and Research
College of Education
University of South of Florida

Major Professor: John M. Ferron, Ph.D.
Jeffrey D. Kromrey, Ph.D.
Eun Sook Kim, Ph.D.
Julia Ogg, Ph.D.

Date of Approval:
May 27, 2014

Keywords: Hierarchical, Multilevel, Synthesis, Single-subject, N=1

DEDICATION

To my Lord and Savior: When no one else was there during those long nights that I wanted to just give up, you never gave up on me! Thank You for giving me strength when I didn't think that I could make it, peace in the midst of the chaos that sometimes surrounded me, and joy when all I really wanted to do was cry.

In creole:

Pou Seyè ak Sovè m: lè pa t gen yon lòt te la pandan nwit long sa yo ke mwen te vle jis bay moute, Ou pa janm te moute sou m! Mèsi pou bay m fòs lè mwen pa t panse m te ka fè l, kè poze nan mitan an nan dezòd la ke pafwa sènen m', ak kè kontan lè sèl sa m te vle fè se kriye.

To my family: for never giving up on me and truly always having my back. This is just a small piece that we can now add to our legacy and history. For all of those times, that I never said it…Thank you! I will be eternally grateful for your love and support throughout this journey.

In creole:

Pou fanmi m: paske nou pa janm bay moute sou mwen epi se verite nou toujou la pou mwen. Sa a se jis yon ti moso nou kapab kounye a ajoute nan eritaj nou an ak istwa nou. Pou tout fwa yo mwen pa janm te di nou ... Mèsi! Mwen pral etènèlman rekonesan pou lanmou ak sipò nou nan tout vwayaj sa a.

ACKNOWLEDGMENTS

I would like to thank my professors for serving as true mentors throughout this process. I would like to especially thank Dr. Ferron for your unwavering support and patience. I could not have asked for anything more in a major professor; to Dr. Kromrey for pushing me to make my document as detailed and precise as possible; to Dr. Kim for removing the limits on my paper and revealing the boundless outlets for my project; to Dr. Ogg for making my paper accessible to the researchers in the field who are dealing with single-case interventions daily.

Furthermore, I would like to thank my fellow graduate students for bringing joy, fun, and laughter to a very tough journey. Last but definitely not least, to my new coworkers at Pinellas County Schools...thank you for your support and encouragement as I finished up this last mile. The timing could not have been more perfect. Thank you for serving as invaluable resources as I make the transition to a new job, a new community, and a new life for me and my family!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Single-case interventions allow for the repeated measurement of a case or participant across multiple time points, to assess the treatment's effect on one specific case or participant. The basic interrupted time series design includes two phases: baseline and treatment. Raudenbush and Byrk (2002) demonstrated that a meta-analysis of large group designs can be seen as a special case of multi-level analysis with participants (level-one) nested within studies (level-two). Raw data from a set of single case design studies have a similar structure. Van den Noortgate and Onghena (2003) illustrated the use of a two-level model to analyze data in primary single-case studies. In 2008, Van den Noortgate and Onghena later proposed that if raw data from several single case designs are used in a meta-analysis, scores can be varied at each of the three levels: over occasions (level-one), across participants from the same study (level-two), and across studies (level-three).

The multi-level approach allows for a large degree of flexibility in modeling the data (Goldstein & Yang, 2000; Hox & de Leeuw, 1997). Researchers can make various methodological decisions when specifying the model to approximate the data. Those decisions are critical since parameters can be biased if the statistical model is not correctly specified. The first of these decisions is how to model the level-one error structure--is it correlated or uncorrelated? Recently, the investigation of the Van den Noortgate and Onghena's (2008) three-level meta-analytic model has increased and shown promising results (Owens & Ferron, 2011; Ugille, Moeyaert, Beretvas, Ferron, & Van den Noortgate, 2012). These studies have shown the fixed effects tend to be unbiased and the variance components have been problematic across a

range of conditions. Based on a thorough literature review, no one has looked at the model in relation to the use of fit indices or log likelihood tests to select an appropriate level-one error structure.

The purpose of the study was two-fold: 1) to determine the extent to which the various fit indices can correctly identify the level-one covariance structure; and 2) to investigate the effect of various forms of misspecification of the level-one error structure when using a three-level meta-analytic single-case model. This study used Monte Carlo simulation methods to address the aforementioned research questions. Multiple design, data, and analysis factors were manipulated in this study. The study used a 2x2x2x2x2x5x7 factorial design. Seven experimental variables were manipulated in this study: 1) The number of primary studies per meta-analysis (10 and 30); 2) The number of participants per primary study (4 and 8); 3)The series length per participant (10 and 20); 4)Variances of the error terms (most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at the upper levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]); 5) The levels for the fixed effects (0, 2 [corresponding to the shift in level]; and 0, 0.2[corresponding to the shift in slope]) 6)Various types of covariance structures were used for data generation (ID, AR(1), and ARMA (1,1); and 7) The form of model specification [i.e. ID, AR(1), ARMA (1,1)], and error structure selected by AIC, AICC, BIC, and the LRT.

The results of this study found that the fixed effects tend to mostly be unbiased, however, the variance components were extremely biased with particular design factors. The study also concluded that the use of fit indices to select the correct level-1 structure was appropriate for certain error structures. The accuracy of the fit indices tend to increase for the simpler level-one error structures. There were multiple implications for the applied single-case researcher, for the meta-analyst, and for the methodologist. Future research included investigating different estimation methods, such as Bayesian approach, to improve the estimates of the variance components and coupling multiple violations of the error structures, such as non-normality at levels two and three.

CHAPTER ONE: INTRODUCTION

Single-Case Designs

Single-case interventions allow for the repeated measurement of a case or participant across multiple time points, to assess the treatment's effect on one specific case or participant. The basic interrupted time series design includes two phases: baseline and treatment. The baseline (pretreatment) phase consists of a series of observations preceding the introduction of a treatment. The baseline phase serves two primary functions: 1) to describe the existing level of performance that is to be altered, and 2) to serve as the basis for which predictions can be made for the participant if the intervention had not been introduced. The treatment phase consists of a series of observations following the introduction of a treatment. Inferences about the research are usually made about the effects of the intervention by comparing different conditions (baseline vs. treatment) presented to the same participant or many participants over time (Kazdin, 2011). There are many commonly used single case designs. The most commonly used design is the multiple-baseline, which includes time-series data from multiple participants (or behaviors or settings) where an intervention is staggered to occur at different time points within the various series (Ferron et al., 2009).

Repeated measures design is based on continuous observations over time for the same subject. This feature of single case research is one of the strengths of this design given that it can allow a researcher to analyze a particular case in depth. However, this can also present challenges in terms of choosing an appropriate data analysis method. The need to model serial dependency, the amount of dependence is typically characterized by the correlation between adjacent time points, had been a great discussion in the literature. Specifically, whether or not single-case data can show serial dependence due to small sample sizes, or how to best estimate the autoregressive parameters to ensure that they are unbiased due to the small *n* was debated in the single case literature (Matyas & Greenwood, 1996; Huitema & McKean, 1991). Nevertheless, studies have shown that there is indeed some correlation beyond random chance in repeated measures design for observations within a single subject (Kratochwill et al., 1974). Barlow, Nock, and Hersen (2009) concluded that autocorrelation may or may not exist given the above debate, however, based upon past research, it would appear reasonable for single-case analysts to examine their data for the presence of autocorrelation. If autocorrelation is assumed to be present in the population, then choosing a method that is appropriate for their data seems ideal.

There are numerous options or tools when analyzing single-case designs. Examples of these analyses options include visual analysis, randomization tests, and multi-level modeling. Additionally, there are a variety of effect size indices that are used to supplement these analyses options. These include non-parametric effect size indices, such as percentage of non-overlapping data (Scruggs & Mastropieri, 1998), a change in R^2 (Allison & Gorman, 1993; Beretvas & Chung, 2008; Kromrey & Foster-Johnson, 1996), or the use of standardized coefficients when applying multi-level models (Van den Noortgate & Onghena, 2003, 2007, 2008). These effect size measures are often used to characterize the size of the intervention effect. Researchers are not only interested in the intervention effect within a particular study, but may also want to know about the intervention effect across studies.

Meta-analytic procedures allow researchers to quantitatively synthesize past research results, and provide evidence for best practices (Hedges & Olkin, 1985). However, there has been no consensus on the best way to synthesize these data. Beretvas and Chung (2008) conducted a narrative review of 25 single-case meta-analyses and found that most of the SSED (single-subject experimental designs) meta-analysts were using non-parametric approaches. They found that meta-analysts were using the simplest indicators as effect size measures, such as change in R^2 (Mooney, Ryan, Uhing, Reid, & Epstein, 2005; Swanson & Sachse-Lee, 2000); and non-parametric methods such as the percentage of non-overlapping data (PND) (Templeton, Neel, & Blood, 2008; Xin & Jitendra, 1999) and/or the percent of all non-overlapping data (PAND) (Bellini, Peters, Benner, & Hopf, 2007; Codding, Burns, & Lukito, 2011) for conducting meta-analysis involving single-case research. One major limitation is that these approaches do not allow for inferences about the treatment effects.

Raudenbush and Byrk (2002) demonstrated that a meta-analysis of large group designs can be seen as a special case of multi-level analysis with participants (level-one) nested within studies (level-two). Raw data from a set of single case design studies have a similar structure. Van den Noortgate and Onghena (2003) illustrated the use of a two-level model to analyze data in primary single-case studies. In 2008, Van den Noortgate and Onghena later proposed that if raw data from several single case designs are used in a meta-analysis, scores can be varied at each of the three levels: over occasions (level-one), across participants from the same study (level-two), and across studies (level-three). Equation 1 below describes the variation within participants that occurs when treatment conditions are compared with a baseline condition (level-one). At the second level, the variation over participants is shown using two regression equations (Equations 2 and 3). Finally, the last set of equations describes the variation across the studies (Equations 4 and 5) that are included in the meta-analysis.

Equations 1-5 below denote the model used to represent the fixed effects and the variance components at each of the three levels. The variable phase is a dichotomous variable representing the baseline phase (phase = 0) and the treatment phase (phase = 1). It should be noted that errors on all three levels are typically assumed to be normally distributed and have a mean of 0. The model is presented below:

Level 1 Equation:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} * phase + e_{ijk}, \quad e_{ijk} \sim N(0, \Sigma_e) \tag{1}$$

Level 2 Equations:

$$\beta_{0jk} = \theta_{00k} + u_{0jk} \qquad \begin{bmatrix} u_{0jk} \\ u_{1jk} \end{bmatrix} \sim N(0, \Sigma_u) \tag{2}$$

$$\beta_{1jk} = \theta_{10k} + u_{1jk} \tag{3}$$

Level 3 Equations:

$$\theta_{00k} = \square_{000} + v_{00k} \qquad \begin{bmatrix} v_{00k} \\ v_{10k} \end{bmatrix} \sim N(0, \Sigma_v) \tag{4}$$

$$\theta_{10k} = \square_{100} + v_{10k} \tag{5}$$

The multi-level approach allows for a large degree of flexibility in modeling the data (Goldstein & Yang, 2000; Hox & de Leeuw, 1997). Researchers can make various methodological decisions when specifying the model to approximate the data. Those decisions are critical since parameters can be biased if the statistical model is not correctly specified. The first of these decisions is how to model the level-one error structure--is it correlated or uncorrelated? The errors in the first-level of the model represent the discrepancy between the values of the outcome observed and of the individuals' growth trajectory (Ferron, Dailey, & Yi, 2002).

There are several options when dealing with the level-one error structure. These options range from assuming that the error structure is uncorrelated, $\sigma^2 \mathbf{I}$ to choosing an appropriate correlated error structure. A method of handling the level-one error structure is to ignore the correlated error structure, subsequently making incorrect assumptions, such as the assumption of independence (Littell, Pendergast, & Natarajan, 2000). A researcher has to decide if the level-one error structure (Kesselman, Algina, Kowalchuk, & Wolfinger, 1999) should be correlated or uncorrelated. If the error structure is correlated, then which structure is best? Is it a first-order autoregressive or a moving average autoregressive model? The most commonly used level-one error structure is $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, which is used when a researcher has decided that the errors are uncorrelated. For correlated level-one error structure, the most commonly used option is the first-order autoregressive, AR (1) (Ferron, Dailey, & Yi, 2002). The use of fit indices to determine the correct level-one error structure has not been examined in the single case literature. Furthermore, the consequences of various forms of misspecification of the level-one error structure have also not been investigated in terms of the meta-analysis of single-case data. Therefore, the study utilized research related to the broader repeated measures or longitudinal literature to design the conditions that were used.

Autocorrelation and Longitudinal Designs

Whether or not to model autocorrelation had been a huge discussion and studied extensively in growth curve modeling (GC modeling) or longitudinal data analysis (Kesselman, Littell, & Sivo, 2003). Growth curve modeling (or longitudinal) data's defining characteristic is that individuals are measured repeatedly over time enabling direct study of change (Diggle, Heagerty, Liang, & Zeger, 2002). A question that commonly arises with both single case and longitudinal designs is whether or not the model has the correct level-one error structure specification, and if so, what is the correct error structure? Many researchers who use uncorrelated error structures commonly assume that $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ (Bryk & Raudenbush, 2002).This commonly used approach should lead the researcher to question or to ask whether or not **Σ** has been misspecified (Kesselman et al 1999; Kwok, West, Green, 2007). Simply assuming that the level-one errors are uncorrelated has shown to lead to inflated Type I errors and biased confidence interval coverage in single case designs (Ferron, Bell, Hess, Rendina-Gibioff, & Hibbard, 2009) and in longitudinal data analysis (Kwok et al., 2007) if in fact, autocorrelation is present in the population.

Conversely, once researchers have decided to model a correlated level-one error structure, there are two commonly used approaches to select an appropriate level-one covariance structure when using multi-level models. Some researchers may choose to specify their model using a simple correlated error structure a priori **(**Kwok, West, & Green, 2007; Murphy & Pituch, 2009). Another method is to rely on fit indices or log likelihood tests to identify the correct covariance structure (Ferron et al., 2002; Kesselman et al., 1999). The study investigated both of these methods and their application to single-case data.

Problem Statement

Single-case research has traditionally been left out of meta-analytic studies, due to the lack of agreement on the best way to meta-analyze single case data (Faith, Allison, Gorman, 1996; Van den Noortgate & Onghena, 2008). Meta-analysis of single case designs would not only allow for the understanding of generalizability, or the treatment effect across studies, but it also affords researchers the benefit of understanding how the treatment's effect relates to specific individuals within a particular study. Meta-analyses generally have three goals. 1) Meta-analytic studies strive to provide a point estimate of the average effect size, in short, a quantitative summary. 2) Meta-analyses strive to provide confidence intervals in which the "true" population effect size is likely to be found. The confidence interval can then aid in the decision as to whether the effect size is significantly different from zero. 3) Meta-analytic techniques can help the researcher search for variables, or moderators, that could help explain the differences or variability among effect sizes. This is the case in which there is a substantial variability among the effect sizes.

Recently, the investigation of the Van den Noortgate and Onghena's (2008) three-level meta-analytic model has increased and shown promising results (Owens & Ferron, 2011; Ugille, Moeyaert, Beretvas, Ferron, & Van den Noortgate, 2012). These studies have shown the fixed effects tend to be unbiased and the variance components have been problematic across a range of conditions. Based on a thorough literature review, no one to date has looked at the model in relation to the use of fit indices or log likelihood tests to select an appropriate level-one error structure. Furthermore, no one has looked at the consequences of general misspecification of the level-one error structure when meta-analyzing single case raw data using the three-level model. Therefore, it is necessary to investigate whether the work that was conducted in the broader repeated measures designs can be applied to smaller samples designs such as single case research.

Study's Purpose

There have been a multitude of articles analyzing fit indices and properly identifying the correct covariance structures in terms of the broader longitudinal area, or growth curve models. A thorough literature search has produced no studies to date looking at fit indices in terms of single case research using multi-level models. Moreover, there has been no research on the consequences of different forms of specification of the level-one error structure when using a three-level meta-analytic single-case model.

The purpose of the study was two-fold: 1) to determine the extent to which the various fit indices can correctly identify the level-one covariance structure; and 2) to investigate the effect of various forms of misspecification of the level-one error structure when using a three-level meta-analytic single-case model. The research questions of interest are as follows:

Research Questions

1. To what extent do fit indices (AIC, adjusted AIC, BIC, log likelihood ratio test) correctly identify level-one covariance structure when using a three-level meta-analytic single-case model?
2. To what extent are the **fixed effect** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?
3. To what extent are **confidence interval width and coverage for the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?
4. To what extent are the **Type I error and power for the test of the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?

5. To what extent are the **variance component** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, and covariance structures, level of the treatment effect), and analysis factors (form of specification)?
6. To what extent are **confidence interval width and coverage for the variance components** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?

Overview of the Study

This study used Monte Carlo simulation methods to address the aforementioned research questions. Multiple design, data, and analysis factors were manipulated in this study. The study used a 2x2x2x2x2x5x7 factorial design. The conditions are briefly introduced here, but were explained with great detail later in Chapter Three. Seven experimental variables were manipulated in this study. 1) The number of primary studies per meta-analysis (10 and 30); 2) The number of participants per primary study (4 and 8); 3)The series length per participant (10 and 20); 4)Variances of the error terms (most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at the upper levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]); 5) The levels for the fixed effects (0, 2 [corresponding to the shift in level]; and 0, 0.2[corresponding to the shift in slope]) 6)Various types of covariance structures were used for data generation (ID, AR(1), and ARMA (1,1); and 7) The form of model specification [i.e. ID, AR(1), ARMA (1,1)], and error structure selected by AIC, AICC, BIC, and the LRT. For each of the 160 data and design conditions, 5000 simulated data sets were generated using SAS IML (SAS Institute, Inc., 2008). These data sets were then specified using the a priori model selection of the level-one error structure and the use of fit criteria or post hoc model selection of the level-one error structure. This study examined the fixed effects, (i.e., the overall average baseline level, the overall average treatment effect, the overall average baseline slope, and the overall average difference between baseline and treatment slope) and the variance components (e.g., the between-person within-study variance in the average baseline level, the between-person within-study variance in the average treatment effect, the between-person within-study variance in the average baseline slope, the between-person within-study variance in the average difference between baseline and treatment slope, the between-study variance in the average baseline level, the between-study variance in the overall average treatment effect, the between-study variance in the overall average baseline slope, and the between-study variance in the average difference between the baseline and the treatment slopes) in a three level multi-level model.

Significance of Study

This study contributed to the ongoing debate of autocorrelation and single-case designs. Moreover, this study provided another method, or opportunity, to include single-case designs in meta-analyses. For methodologists, this study can serve to demonstrate the importance of selecting the correct level-one error structure, and how the error structure can impact the parameter estimates and inferences made from those estimates. For the applied researcher and practitioner, this study can serve to illustrate how fit indices can be used to select the correct level-one error structure. Additionally, this study can serve to demonstrate the difference between selecting the error structure a priori or using fit indices, and the impact of the correct level-one error structure on the parameter estimates obtained from the model. This study serves also in terms of the some design features that can better allow for the meta-analysis of single-case designs. Some of these issues include sample size on all three levels and baseline stability.

The conditions in this study serve to not only replicate and extend previous research in the methodological area. The conditions were also chosen to represent current applied works in this area. Specifically, the conditions that were used in this study were drawn from a combination of methodological works and applied meta-analysis that were done using multi-level models.

Several aspects of this study distinguish it from previous works that have investigated the three-level model to meta-analyze single-case data. One primary aspect is the appropriate use of fit indices to determine the correct model specification when dealing with small samples, i.e. single case designs. Additionally, the consequences of misspecifications of the level-one error structure were examined when meta-analyzing single case research.

Limitations

The data in this study were simulated based on specific design conditions. Those conditions were chosen based on a review of single-case literature, meta-analyses of single-case data, and applied work that was done using the three-level model to aggregate data across studies. The specific conditions chosen for this study are only a portion of the possible options that could have been included in this study. Therefore, the results of this study can only be generalized to studies with similar conditions. Any conclusions beyond the observed conditions should be interpreted with caution.

Table 1

Design Factors and Level for each Factor

				Error Variances					
				Most Variance at Level-1: [σ^2=1; Σ_u = .5, 0.05, .5, 0.05; Σ_v = .5, 0.05, .5, 0.05]			Most variance at the upper levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]		
				Data Generation					
				σ = ID	σ = AR(1)	σ = ARMA (1,1)	σ = ID	σ = AR(1)	σ = ARMA (1,1)
				Model Specification					
				σ = ID σ = AR(1) σ = ARMA (1,1)	σ = ID σ = AR(1) σ = ARMA (1,1)	σ = ID σ = AR(1) σ = ARMA (1,1)	σ = ID σ = AR(1) σ = ARMA (1,1)	σ = ID σ = AR(1) σ = ARMA (1,1)	σ = ID σ = AR(1) σ = ARMA (1,1)
Number of Participants per Meta-Analysis	Number of Participants	Number of Observations	Level for the Fixed Effect						
10	4	10	0						
			2						
		20	0						
			2						
	8	10	0						
			2						
		20	0						
			2						
30	4	10	0						

Table 1 (Continued)

Design Factors and Level for each Factor

Number of Participants	Number of Observations	Level for the Fixed Effect
		2
	20	0
		2
8	10	0
		2
	20	0
		2

Definition of Terms

Autocorrelation. The extent to which the values of the observed behavior at time t (Y_t) are correlated with values at t – i, or Y ($_{t-i}$) (Matyas & Greenwood, 1996).

Bias. The difference between a known parameter and an expected parameter estimate, E(□).

Confidence interval coverage. The proportion of 95% confidence intervals that contain the estimated parameter. This outcome was aggregated across replications within each condition to represent the average confidence interval coverage.

Confidence interval width. The difference between the upper and lower limits of the 95% confidence intervals for the estimated parameter. This outcome was aggregated across replications within each condition to represent the average confidence interval width.

Effect size. A measure of the magnitude of the relationship between two variables.

Fit Indices. Akaike Information Criterion (AIC, closer to zero), Adjusted Akaike Information Criterion (AICC, closer to zero), Bayesian Information Criterion (BIC, closer to zero), and Log likelihood ratio test (LRT, statistically significant at $\alpha = .05$). These indices will determine the best model to approximate the data.

Fixed effects. Parameter estimates of the coefficients represented in the multi-level model [e.g. overall baseline level, overall average treatment effect (shift in level), overall baseline slope, and overall treatment effect for the slopes (difference in slopes)]

Hierarchical Linear Modeling (HLM). This term is commonly referred to as multi-level modeling. Multi-level modeling can include two levels: 1) a level-one submodel that describes an individual's change over time; and 2) a level-two model that describes how these changes vary across individuals. Together, these two levels of equations represent a multi-level statistical model (Raudenbush & Bryk, 2002). This technique is useful when dealing with nested data.

Kenward-Roger Degrees of freedom method. This method was developed as an extension of the Satterthwaite method to approximate the degrees of freedom; it adjusts for small sample sizes and works well with complex variance structures (Ferron, Bell, Hess, & Hibbard, 2009).

Mean-Square Error. A measure of the average squares of error.

Meta-analysis. The quantitative synthesis of study results that involves combining study outcomes across studies to evaluate and summarize research findings.

Non overlap of all pairs (NAP). This new index summarizes data overlap between each of the data points in Phase A and each of the data points in phase B.

Over-specification. This is a form of misspecification, but explicitly involves the model that specifies a more complex level-one error structure than the level-one error structure of the data (e.g. a model that specifies σ = AR(1), when the true level -1 error structure σ **= ID**).

Percentage Exceeding the Median (PEM). Describes the percentage of phase B (intervention) data points exceeding the median of the A phase (baseline).

Percentage of All Non-Overlapping Data (PAND). This is defined as "percent of all data remaining after removing the minimum number of data points which would eliminate all data overlap between phases A and B.

Percentage of Non-Overlapping Data (PND). The percentage of phase B (treatment) data points which exceed the single highest phase A (baseline) datum point (or below if the lowest point of data points in the baseline phase if the undesirable outcome or behavior is expected to decrease) (Scruggs & Mastropieri, 1998).

Primary Studies. The original studies that comprise the sample for the meta-analysis.

Randomization design. Refers to the presentation of alternative interventions in random order, usually with the restriction that the conditions are presented an equal number of times (Kazdin, 2011).

Satterthwaite degrees of freedom method. A method to approximate the degrees of freedom that was developed to be used with unbalanced designs and complex error structures.

Series length. The level-one sample size for the participants in single-case research. This is also referred to as the number of observations or measurements.

Single-case research. The repeated measurement of a case or participant across multiple time points, to assess the treatment's effect on one specific case or participant.

Treatment effect. The change in the outcome variable as a response to being in the intervention phase. This can refer to the *change in level* or the *change in slopes*.

Under-specification. This is a form of misspecification, but explicitly this involves the model that specifies a simpler level-one error structure than the level-one error structure of the data [e.g. a model that specifies σ **= ID** when the true level-one error structure is σ = AR(1)].

Variance components. The parameters that estimate the variation within person, between persons within studies, and between studies.

CHAPTER TWO: LITERATURE REVIEW

This literature review was divided into four parts. First, a brief overview of single-case designs was provided. Secondly, the analysis techniques for primary single case studies were described. Third, the various methods for synthesizing single case data across studies, particularly using multi-level modeling were discussed. Lastly, the examination of the selection of the level-one error structure: either through the use of fit criteria, selecting level-one error structure a priori, or conducting a sensitivity analysis.

Single-Case Designs

In large group studies, focus is generally placed on the average amount of change across groups from pre-treatment (or pre-intervention) to post-treatment. Focusing on this type of change can cause one to miss the opportunity to understand how, why, and when such a change has occurred (Barlow, Nock, & Hersen 2009). Single-case interventions allow for the repeated measurement of a case or participant across multiple time points, or in other words, to assess the treatment's effect on one specific case or participant. The most fundamental design element of single-case research is the reliance of repeated observations or measurements of performance over time for each participant. The basic interrupted time series design includes two phases: the baseline and treatment phases. The baseline (pretreatment) phase consists of a series of observations preceding the introduction of a treatment. The baseline phase serves two primary functions: 1) to describe the existing level of performance that is to be altered; and 2) to serve as the basis for which predictions can be made for the participant if the intervention had not been introduced. The treatment phase also consists of a series of observations following the introduction of a treatment. Inferences about the research are usually made about the effects of the intervention by comparing the different conditions (baseline vs. treatment) presented to the same participant or many participants over time (Kazdin, 2011).

Design Characteristics

There are many designs that can be used in single case research. These designs influence or attempt to reduce the internal validity threats that can be present in this type of research. These designs include, but are not limited to, the basic AB design, the repeated ABAB design, the alternating treatment design, and the multiple baseline designs (Barlow et al., 2009; Kazdin, 2011). Kazdin (2011) used the ABAB design to illustrate the key elements present in many of the commonly used single-case designs. The simple ABAB design examines the effect of an intervention by alternating the baseline condition, which is referred to as the A phase, and the treatment condition, also known as the B phase. The A and B phases are then repeated again. Ideally, one would observe that the behavior increased (given that is what the researcher hypothesized) when the participant(s) were in the B phase and then returns back to its original baseline levels once the intervention has been withdrawn or removed. Finally, the performance again increases in the last treatment, or B phase (Kazdin, 2011). This commonly used design attempts to control for various threats to internal validity. More specifically, looking at the intervention and then withdrawing the intervention and adding the intervention again reduces the likelihood that an external event could have caused the observed changes. The design only leaves one plausible explanation--the intervention caused the observed change. Figure 1 shows an example of an ABAB design where the author's purpose was to examine the relationship between social stories and undesirable behavior. More specifically, the researcher sought to investigate if social stories could reduce undesirable behaviors in a student with autism. The figure displays the first and third phases as the baseline phases, followed by the second and fourth phases (intervention phases). During the intervention, both social stories were read for the student to remind him of appropriate social behaviors (Figure 1 adapted from Lorimer & Simpson, 2002). The ABAB design characteristics are similar across many single case designs.

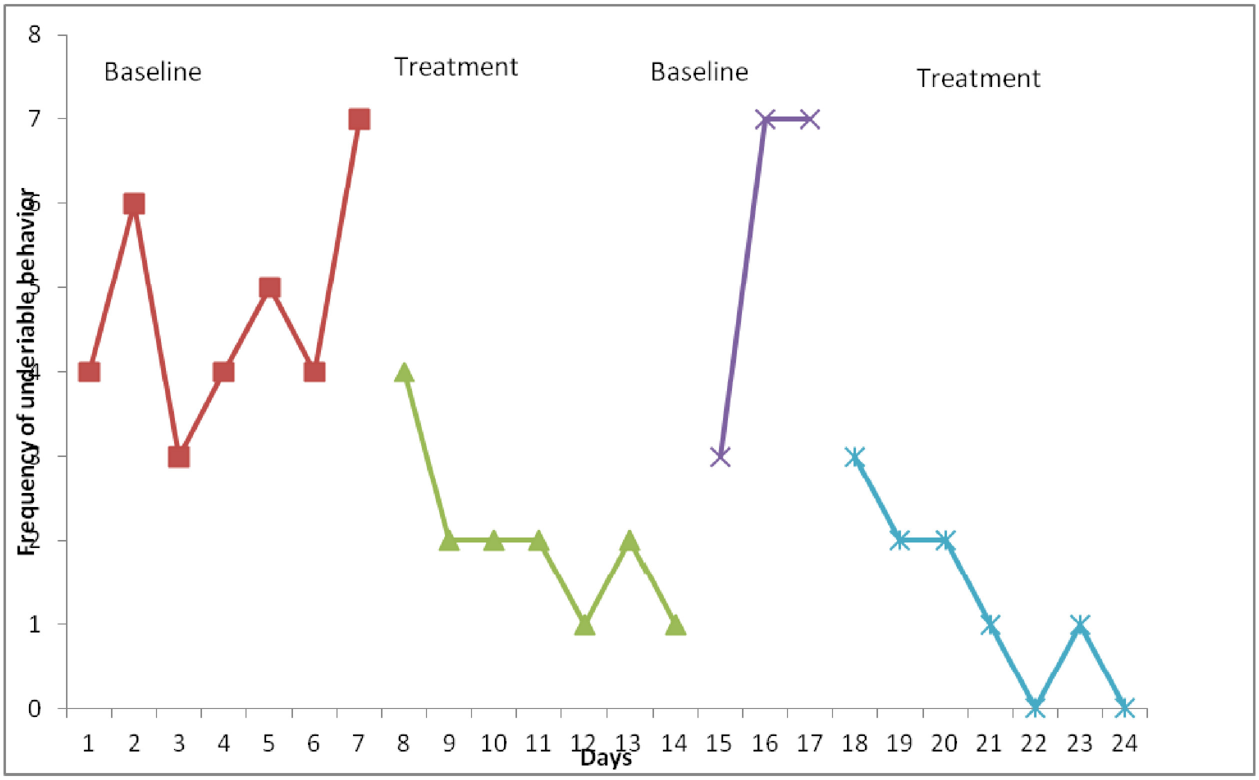

Figure 1. An example of an ABAB graph for an undesirable behavior. This figure illustrates the ABAB design with the use of social stories to reduce the undesirable behavior.

Another complex design involves one phase change, from A phase to B phase, across multiple participants or settings. This single case design, called the multiple-baseline, includes time-series data from multiple participants (or behaviors, or settings) where an intervention is staggered to occur at different time points within the various time series (Ferron, Bell, Hess, & Hibbard, 2009). This allows for several cases or settings to be analyzed simultaneously within a study. Additionally, Kazdin (2011) describes multiple-baseline designs explaining that the effects are demonstrated by introducing the intervention to different baselines, whether that is behaviors or persons. If each baseline changes, and the expected outcome occur, each time the intervention occurs then we can feel more comfortable with attributing the effect to the intervention, rather than to extraneous factors. Multiple-baseline designs do not share the practical or ethical issues with some of the other designs, such as the ABAB design. The intervention does not need to be withdrawn once the intervention is introduced to a particular baseline.

The power of these designs is illustrating that the expected change occurs only when the treatment or intervention is directed at the behavior, setting, or subject (Barlow et al., 2009). Figure 2 (adapted from Ferron et al., 2009) below illustrates a graph of a multiple baseline design. The multiple-baseline design is the most commonly used design in the single case research. Shadish and Sullivan (2011) located, digitized, and coded 809 single case designs from 113 studies in 2008 in 21 different journals in a variety of fields including psychology and education. They found that the majority of single-case designs included some form of multiple baseline design, either alone or in combination with another design. Moreover, approximately 79% of the single case studies included some form of multiple baseline design.

Analysis Alternatives

There are several methods for analyzing single-case data. The next few sections will review a portion of the commonly used methods for the analysis of single-case research.

Visual Inspection or Analysis. Kazdin (2011) refers to visual inspection as reaching a judgment about the reliability or consistency of intervention effects by visually examining the data. The author goes on to describe the experimental criterion as a comparison of performance during the intervention with what the performance would look like if the intervention had not been implemented. Visual analysis has been the primary analysis for single case data (Busk and Marascuilo, 1988; Fisch, 2001; Kazdin, 2011).

Figure 2. Multiple-Baseline Design adapted from Ferron, Bell, Hess, & Hibbard, 2009

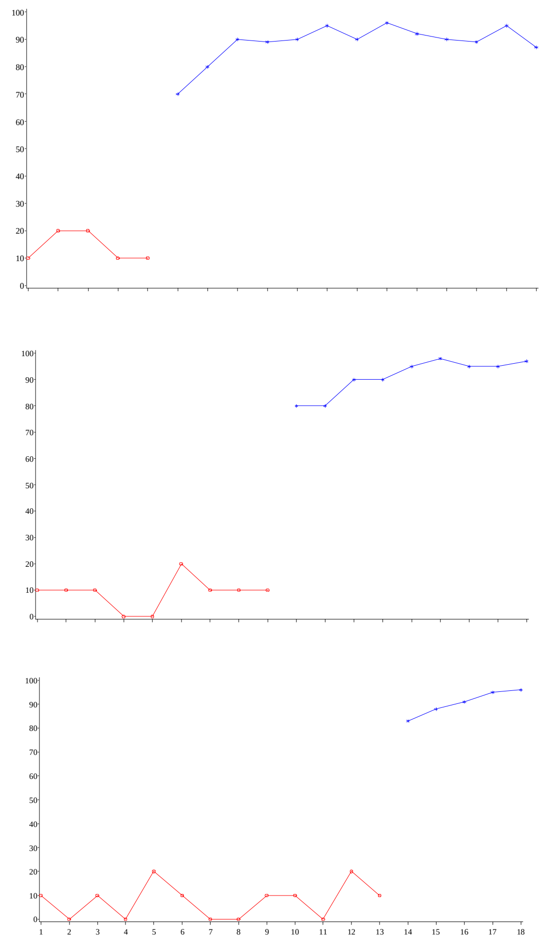

Figure 2. Visual graphs 3 participants in a multiple-baseline study. Notice that the 3 participants have different lengths in terms of baselines and treatments.

In a survey and review of single case literature, Busk and Marascuilo (1988) found that visual analysis was still the dominant form of analysis used in single case research designs, and furthermore Kazdin (2011) recently describes visual inspection as the primary method of data evaluation in single-case research. Proponents of visual analysis argue that visual analysis have low Type-I error rates (Brossart Parker, Olson, & Mahadevan, 2006; Gorman & Allison, 1996). There has been a reluctance to the sole use of visual analysis because the lack of inter-rater reliability or the high incidence of judges' error that has been seen in several studies. Past research had found that reliabilities tend to be low to moderate (Brossart et al., 2006). Brossart et al. (2006) conducted a study to test the inter-rater reliability among 15 experienced judges in determining the extent of the intervention effect. They found that their inter-rater reliability was very poor, unless the intervention effect was substantial, which could be easily identified by the visual analyst. Furthermore, low reliabilities and high response bias has been noticed increasingly when the intervention effect involves trends (Fisch, 2001). The author conducted a review of a multitude of studies that investigated visual analysts' capabilities of determining a treatment effect in the presence of complex data, such as when treatment effect involved both shift in level and shift in trends, in the presence of autocorrelated errors. Fisch (2001) concluded that the most striking feature across all of the studies that were reviewed is the participants' inability to detect trends when they were present alone or in combination with a shift in level treatment effect. This finding supported the low inter-rater reliability found in past literature. This low inter-rater reliability, more specifically when the data are complex, such as in the presence of autocorrelated errors, have been the cause of many disputes among researchers who have argued that visual analysis is not enough to detect intervention effect (Brossart et al., 2006; Huitema, 1986; Kazdin, 2011).

Knowing that serial dependence may be present in single case data, visual analysis alone may not be sufficient. Visual analysis and inspection cannot account for the underlying trend or pattern that may exist within the data due to autocorrelation (Huitema, 1997; Kazdin, 2011;

Matyas & Greenwood, 1990). Likewise, visual analysis cannot discern whether an intervention is effective above and beyond some underlying pattern that would exist in the presence of autocorrelation. In the aforementioned study, Matyas and Greenwood (1990), the aim was to examine the false alarms and miss rates in simple designs, such as AB panels, using visual analysis. Serial dependence in the time series was also systematically varied to determine its effect on rater reliability. A sample of 37 graduate students were asked to make judgments (void of intervention effect, a level change, a trend change, combined level and trend change, or other type of systematic change during intervention) on 27 charts. Variations in false alarms had a significant interaction effect with degree of serial dependence. Conversely, this was not the case with miss rates. Most of the miss rates were below 10%. This study contributed to the idea that the conservatism by visual analysts may be compromised in the presence of serial dependent data. In a much earlier study, Jones, Weinrott, and Vaught (1978) asked 11 experienced judges to visually analyze 24 graphs from a reputable journal. The majority (83%) of the graphs had statistically significant (p <.05) lag-1 autocorrelation that ranged from .40 to .93. The judges had to determine if a "meaningful" shift in level had occurred. Meaningful solely referring to the reliability of change. The authors found the agreement of visual judges with times series analysis would be lowered when serial dependence existed in the data and when statistically significant results were found by the times series analysis. This is shown to be problematic given the fact that autocorrelation or serial dependence is likely to exist in single-case designs and one would hope that the shift in level or treatment effect is statistically significant.

Visual judges may have trouble identifying intervention effects other than shifts in level or changes in linear trend. These issues seem even more prominent when short series lengths are introduced, as is common with single case designs (Matyas & Greenwood, 1996). The difficulty in determining whether or not there is a treatment effect using visual analysis continues to be an important concern in the literature, and this concern is magnified in the presence of correlated errors.

Matyas and Greenwood (1990) explicated that visual analysis tends to become too liberal when positive autocorrelation is present, thereby increasing our Type I errors. Negative autocorrelation tends to produce more conservative values for visual judgment. Finally, the current verdict is that serial dependence or autocorrelation may exist and should be taken into account when evaluating single-case data (Kazdin, 2011). Given this conclusion, and the issues that correlated errors may pose to the visual analyst, it seems reasonable, or almost necessary to supplement visual analysis with some statistical technique (Barlow et al., 2009; Kazdin, 2011).

Overlap Statistics. Another type of analysis that complements the visual analysis is non-regression type indices or effect sizes, sometimes referred to as overlap statistics. This analysis category contains a few options. The percentage of non-overlapping data (PND) effect size measure can be explained as the percentage of phase B (treatment) data points which exceed the single highest phase A (baseline) data point (or below the lowest point of data points in the baseline phase if the undesirable outcome or behavior is expected to decrease). One conceptual advantage is in its meaningfulness to practical researchers; for example, PND scores of over 90 (i.e. 90% of treatment observations exceed the highest baseline observation) (Scruggs & Mastropieri, 1998).

The Percentage of All Non-Overlapping Data (PAND) is defined as the percent of all data remaining after removing the minimum number of data points which would eliminate all data overlap between phases A and B. There are several other non-parametric effect sizes used in the single case literature. PAND has similar features to PND, however, avoiding some of the major criticisms. First, PAND uses all of the data from both the baseline and the treatment phases. More importantly, PAND can be translated to reflect an actual effect size, such as *Pearson's Phi* or *Phi*^{2} (Parker, Hagan-Burke, Vannest, 2007). The percentage exceeding the median (PEM), describes the percentage of phase B (intervention) data points exceeding the median of the A (baseline) phase. The null hypothesis of this approach is if the treatment has no effect, then the points will fluctuate around the middle line. Namely, the points have a 50% chance of being above or below the median of the previous baseline phase (Ma, 2006). The author proposed the PEM as an approach to compensate for some of the shortcomings of PND; one such example is in the presence of ceiling or floor data points in the baseline.

Recently, Parker and Vanest (2009) introduced a fourth measure: non overlap of all pairs (NAP). This new index summarizes data overlap between each of the data points in Phases A and B. A non-overlapping pair will have a treatment (or B phase) data point that is higher than its corresponding baseline (or A phase) data point. They designed this new index to remedy the perceived weaknesses of the other indices. These shortcomings include a) lack of knowledge regarding the underlying distribution, which then makes it difficult to calculate a confidence interval around the effect size (PND); b) a weak relationship between other known effect sizes (PEM); c) low ability to discriminate among published studies (PEM, PND); d) low power for single case designs, which typically have short series length (PND, PAND, PEM); and e) the other indices rely highly on visual analysis, which can lead to human error in hand calculations from the graphs (PND, PAND, PEM) (Parker & Vanest, 2009). The authors found that the NAP was loosely comparable with the previously mentioned effect sizes. The authors concluded that PAND was the strongest index, with the greatest precision and power. Howbeit, none of the indices could discriminate for a large number of samples, particularly among the more successful interventions (Parker & Vanest, 2009). Another common shortcoming among these indices is that the overlap statistics does not handle intervention effects that involve trends.

Randomization Tests. In recent years, there has been increased attention to the use of statistical methods that do not rely on the traditional parametric assumptions (Barlow et al., 2009). One of the greatest advantages to randomization tests is that no assumptions need to be made about the data. For example, it is a distribution-free test statistic. Distribution-free tests that rely solely on the information from the sample have gained increased interest in recent decades (Barlow et al., 2009). The randomization design refers to the presentation of alternative interventions in random order, usually with the restriction that the conditions are presented an equal number of times (Kazdin, 2011). Due to the random assignment of the intervention on any particular day, the results are amenable for several statistical tests (Edgington, 1996).

In single-case designs, a true experiment can be distinguished from a quasi-experiment by the use of random assignment of treatment to measurement occasions (Onghena & Edginton, 2005). There are two types of randomization schemes: alternation randomization and phase randomization. The latter is commonly used in behavioral and educational interventions. The implementation of the phase randomization occurs when the intervention is optimally introduced over the course of several measurement occasions occurring in a predetermined order, or when the order of the introduction of the intervention is predetermined, i.e. baseline phase comes before intervention phase (Onghena & Edginton, 2005).

The logic behind these tests is simple: the null hypothesis is if the intervention has no impact on the observed dependent variable, then the actual observations are not influenced by the intervention, therefore the observed scores simply reflect naturally occurring scores. The data are then analyzed by looking at all possible permutations, or combinations, that could have occurred.

All of the possible outcomes make up the randomization distribution (Barlow et al., 2009). When the order in which the treatments are applied is random, then the design meets the criteria of a randomization design (Kazdin, 2011). Baseline conditions count as a possible treatment phase. Then one can see how randomization can be extended to include the multiple-baseline design. Kratochwill and Levin (2010) conducted a study in which the primary purpose was to provide scientifically credible extensions of various types of single case designs that incorporated randomization. The authors concluded that incorporating some type of randomization in even the most basic type of single case design, such as the AB design could increase the internal validity of the study and allow investigators to draw more valid inferences.

The discussion of randomization tests and power is a significant one in the literature. Ferron and Onghena (1996) estimated the power of randomization tests used with single-case designs involving random assignment of treatments to phases. The authors simulated 120 conditions crossing 6 effect sizes (0, 0.2, 0.5, 0.8, 1.1, and 1.4); 4 levels of autocorrelation (0, 0.3, 0.6, -0.3); and 5 phase lengths (4, 5, 6, 7, and 8). The authors found that estimating power depended not only on the type of design, but there was an interaction effect between type of design and autocorrelation. Positive autocorrelation led to greater power in the random assignment of *treatments to phases* design; while negative autocorrelation had the opposite effect. Based on this study, researchers should explore ways of increasing the power of randomization tests used in conjunction with treatment to phase designs.

Although randomization tests have shown to be efficient, there are limitations. The first of those limitations is related to the statistical power. For the phase design, the power is approximately 10% less than that of an ordinary t-test (Ferron & Onghena, 1996). Phase randomization designs gain increasing power with increasing phase changes; however, the researcher needs to determine how many observations are necessary for each of the phases. Another limitation is there may be instances where the number of observations, outcome measurements, and materials may need to be increased in order to get the correct number of desired comparisons between the AB phases (Kratochwill & Levin, 2010). Randomized designs may also limit the kinds of statistical analysis that are applied to the data. Non-parametric methods are applicable; nonetheless, these tests require a sufficient number of observations or phases to have adequate power to detect an intervention effect (Ferron & Onghena, 1996; Kratochwill & Levin, 2010). Randomization tests can tell us that, yes, there is an effect. However, in order to explore the size or precision of that effect, parametric methods are necessary.

Classical Statistical Modeling. Singer and Willett (2003) described a statistical model as mathematical representations of population behavior. The authors go on and explain that the models describe salient features of the hypothesized process of interest among individuals in the target population. In order to describe these processes and make statements about the populations, statistical models are expressed using parameters, such as intercepts, slopes, and variance components. Gentile, Roden, and Klein (1972) first suggested the use of statistical models, such as t-test or an ANOVA based method for modeling the treatment effect for single case designs. However, even with the simplest design in mind, the AB design, the assumptions of analysis of variance and regression analysis are likely to be violated. The first assumption is normal distributions and equal variances of scores within each level of the independent variable. This may be violated given the treatment effect may alter not only the means, but the distributions of the dependent variable, such as the variance, skewness, and kurtosis (Gorman & Allison, 1996).

Regression based approaches have also been suggested as a possible option to compare the intervention between the baseline and the treatment phases (Huitema & McKean, 1998). Equation 6 represents an outcome (Y_i) that is modeled on time point *i* for each participant. β_0 is the expected score (baseline) for each participant. The expected treatment effect (the difference in means for the baseline and treatment phases) for the participant is represented by $\beta_{1.}$ *Phase* is a dichotomous variable that is coded 0 for baseline and 1 for treatment phase. The within phase error is modeled by e_i (σ^2_e represents the variance of e_i).

$$Y_i = \beta_0 + \beta_1 * phase + e_i \qquad (6)$$

Equation 6 is the most basic model and can be further extended to include terms to evaluate trends in both the baseline and the treatment phases (Center, Skiba, & Casey, 1985-1986; Huitema & McKean, 2000). This is modeled by the following equation:

Regression equation including trends for both phases:

$$Y_i = \beta_0 + \beta_1 * phase + \beta_2 * time + \beta_3 * phase * time + e_i \qquad (7)$$

Again, this model is an extension of the model above, with the difference being now this model controls for time and the interaction of time with the treatment effect. The first two coefficients or fixed effects of the model have similar interpretation as above, but now controlling for time. In other words, the β_1 (phase effect) now represents the level shift at a particular point in time. The phase effect can greatly vary depending on which time point is chosen. Moreover, β_1 is the difference between two predicted values (one for the baseline regression and one for the phase regression) based on all data both before and after intervention. Specifically, the β_1 is the value of $Y_{\text{treatment}}$ predicted at the first point in treatment minus the Y_{baseline} at the same time point, i.e. first time point in treatment; this difference is an estimate of the level change associated with an intervention (Huitema & McKean, 2000). Additionally, $\beta_{2,}$

now represents the baseline's slope, or the trend during the baseline phase; and similarly, $\beta_{3,}$ represents the difference in trend (or slope) between the baseline and the treatment phases at a particular point in time.

Although both the ANOVA and regression type approaches have been suggested as analysis options, the greatest concern is the violation of the assumption of independent residual errors. In single case designs, there would be likely to have some carry over from one time point to the next. For example, students may remember concepts from previous sessions; may feel ill or sick one day that could affect the next several days; or drug interventions may take some time to "wash out". These are examples of what we would call serially dependent data, or autocorrelation which was explained earlier in the chapter. Specifically, statisticians would say that these behaviors, and thus the residuals from the statistical models, are correlated due to successive scores are more similar to each other than would be predicted by chance (Gorman & Allison, 1996). Ostrom (1990) showed that as autocorrelation of residuals increases, computed t-tests for regression weights values may be biased. Moreover, he showed that with modest autocorrelation (> .4) that the observed value would be more than twice the true t-value. When there is positive autocorrelation, the standard errors tend to be smaller, thus leading to more Type I errors or false rejections. On the other hand, when autocorrelation is negative, then the standard errors tend to be larger, thus this will lead to smaller F-values and t-values, causing increasing Type II errors or misses (Kazdin, 2011; Matyas & Greenwood, 1990). Another concern with the use of the traditional t- or F-tests is regarding trends. These parameters are calculated based on means and variances alone, they are strongly discouraged when trends are present in the data (Barlow et al., 2009).

To combat some of the issues related to correlated errors, a generalized least squares (GLS) method was proposed. This method would allow for the flexibility of the regression type approaches, while also controlling for the serial dependency among the data. Maggin et al., (2011) described three general criteria that effect sizes must have: 1) effect sizes must be consistent with the logic of visual analysis; 2) it must control for threats to interpretation, such as autocorrelation and within-phase trends; and lastly, 3) the effect size must have certain statistical properties, i.e. readily interpretable by researchers from a variety of fields. The authors posit that an ideal effect size would provide the flexibility of regression type methods while also modeling autocorrelation.

GLS allows researchers to model autocorrelation, using some basic assumptions that are similar to the assumptions used in other statistical analyses. Effects sizes are derived using a four-step process which includes a model to control for autocorrelation, regression lines estimated for both phases for comparison, GLS regression used to calculate an effect size taking into account the slope and intercept parameters between phases, and finally, the overall effect size utilized in hypothesis testing (Maggin et al., 2011). The author's second purpose was to demonstrate through applied examples that the GLS functioned according to the aforesaid criteria of an ideal effect size measure. Specifically, the authors applied the GLS approach to two published studies that represented strong visual effects that also displayed significant autocorrelation. They found that the GLS effect size did indeed support the visual analyses and the GLS regression method did control for threats to interpretation, such as the presence of autocorrelation. They also found several limitations to the GLS methods. The authors suggested that more work needed to be done in finding the most appropriate method for estimating the autocorrelation parameter. The authors further concluded that their findings led to an opportunity to find appropriate bootstrapping methods that can be added to the GLS method. Additionally, the set of limitations also included the large amount of data, which may not be feasible in single case designs, which are necessary to ensure that the autocorrelation is estimated accurately. Researchers need to be aware of each of their individual data patterns (increasing trend in baseline, immediate level shift, and a decreasing trend as the intervention continues), for there are instances where the parameter estimates could be inaccurate.

The research presented thus far in this chapter has illustrated that autocorrelation can affect the decision made regarding the intervention effect when using both visual and statistical analyses. The key issue still remains regarding correct model identification for the purpose of forecasting and comparing (Matyas and Greenwood, 1996). There are statistical analyses, such as multi-level modeling, that are robust to the violation of independent errors, similar to the GLS method. Additionally, the development of hierarchical linear models or multi-level modeling has created a powerful set of techniques for research on individual change and change across participants, similar to those needed for single case designs (Raudenbush & Bryk, 2002).

Multi-level Modeling. The basic regression model is designed for cross-sectional data, however, in terms of single-case research, or repeated measures data, a model that embodies two types of research questions is necessary. This model addresses the *within-person* change and the *between-person* differences in change (Singer & Willet, 2003, p. 47). More specifically, this suggests that a model for change must include two levels: 1) a level-one sub-model that describes individuals change over time; and 2) a level-two model that describes how these changes vary across individuals. Together, these two levels of equations represent a multi-level statistical model (Raudenbush & Bryk, 2002) for analyzing primary single case studies. This model can further be thought of as an extension of the equations presented earlier. This extension now allows the researcher to synthesize across participants within a particular study. This model is represented and explicated:

Level 1 Equation:

$$Y_{ij} = \beta_{0j} + \beta_{1j}*phase_{ij} + \beta_{2j}*time_{ij} + \beta_{3j}*phase_{ij}*time_{ij}' + e_{ij} \quad (8)$$

Level 2 Equations:

$$\beta_{0j} = \theta_{00} + u_{0j} \quad (9)$$

$$\beta_{1j} = \theta_{10} + u_{1j} \qquad \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix} \sim N(0, \Sigma_u) \quad (10)$$

$$\beta_{2j} = \theta_{20} + u_{2j} \quad (11)$$

$$\beta_{3j} = \theta_{30} + u_{3j} \quad (12)$$

The score on the dependent variable on measurement occasion *i* for subject *j* is Y_{ij}. The *phase* variable is a dummy-coded indicator reflecting whether the observation is in the baseline phase ($phase_{ij}$ = 0) or the intervention phase ($phase_{ij}$ = 1). The coefficient for the $time_{ij}$ variable represents the slope during the baseline phase, $time_{ij}$, and the coefficient for the interaction term, $phase_{ij}*time'_{ij}$, reflects the difference between the baseline and intervention phases' slopes. The *time* variable can be centered in a variety of ways which may be helpful for interpreting the model parameters (Baek, Moeyaert, Petit-Bois, Beretvas, Van den Noortgate, & Ferron, 2013).

Specifically, for this study, the *time* variable is uncentered for the baseline phase's slope such that for the first measurement occasion in the baseline phase, $time_{ij}$= 0. However, the $time_{ij}'$ variable in the interaction term, $phase_{ij}*time'_{ij}$ is coded such that $time'_{ij}$ = 0 for the first observation in the intervention phase. Therefore, the expected score during the baseline phase if extended one observation into treatment is equal to β_{0j}; and the expected score of the treatment phase at this same point in time is β_{1j} higher. As a result, β_{1j} refers to the immediate level change associated with the intervention (Huitema & McKean, 2000). β_{2j} represents the linear trend during the baseline phase, and the linear trend for the treatment phase was β_{3j} higher. In other words, β_{3j} indicates the effect of the intervention on the trend or the shift in slopes. Both β_{1j} and β_{3j} are needed to fully describe the treatment effect. The error in the level-one model (e_{ij}) can be assumed to be normally distributed with some covariance $\sum$. However, a variety of alternative structures for $\sum$ can be assumed, including an independent structure ($\sigma^2\mathbf{I}$) and a first-order autoregressive structure [AR(1)].The level-one equation (Equation 8) is similar to Equation 7 which was previously described, with the exception that now the model is allowed to vary across participants.

It should also be noted that the dependent variable in this study is assumed to be continuous. The use of continuous variables in single-case studies is common in terms of mathematics achievement (Billingsley, Scheuermann, & Webber, 2009) or words read per minute (Tam, Heward, & Heng, 2006). There are various types of outcomes that are commonly used in single case studies, such as binary, ordinal, or count outcomes, for example, counting the number of times that a student talks out without raising their hands or the number of times that a student leaves their seat. These examples would require different types of assumptions using a Poisson distribution (Shadish & Rindskopf, 2007; Shadish et al., 2008).

Many researchers have suggested the use of multi-level modeling to analyze single case data, particularly when correlated errors are present in the data (Ferron, Bell, Hess, & Hibbard, 2009, Raudenbush & Bryk, 2002). The use of multi-level modeling allows for flexibility in handling nesting of observations within a participant, heterogeneous variances, and moderating effects (Ferron et al., 2009; Shadish & Rindskopf, 2007; Van den Noortgate & Onghena, 2003). Multi-level modeling (MLM) estimates of the individual effects are Empirical Bayes (EB) estimates, which depend not only on the data from the individual, but also on the data from other participants. EB estimates are obtained by creating an average of an estimate that is based on information solely from that individual and an estimate that is based on the average of all of the participants' data (Raudenbush & Bryk, 2002). Furthermore, the authors, Raudenbush & Bryk (2002) pointed out that raw data from a set of single case design studies had a similar structure as the two-level meta-analysis for group designs: observations nested within individuals (level one) and across participants within a study (level two).

Van den Noortgate and Onghena (2003) also illustrated the use of a two-level model to analyze data in the individual studies in single case designs. Simulation research has shown promising results in terms of utilizing two-level models to analyze single case data (Ferron, Bell, Hess, Rendina-Gibioff, & Hibbard, 2009). Kwok, West, Green (2005) described many additional benefits to using MLM in the broader context of repeated measures designs. One of the benefits that they cite is the ability to look at moderators and cross-level effects. For example, a researcher may be interested in looking at whether individual characteristics such as age and gender can influence someone's growth or reaction to an intervention. The second advantage, which is more relevant to the present study, is that the covariance matrices of both the between subject random effects and the within subject random errors can be flexibly and simultaneously modeled. This notion of modeling different covariance matrices can be extended to analyzing primary single-case designs and then extended further to the meta-analysis of single case designs.

The need to model autocorrelation has risen in the single case literature and many researchers have conducted Monte Carlo studies addressing the issue of correlated errors when using single case designs (Ferron et al., 2009; Ferron, Farmer, & Owens, 2010). The study conducted by Ferron et al. (2009) examined the interval estimate of the average treatment effect for two methods of specifying level-one error structure ($\sigma^2 \mathbf{I}$ or first-order autoregressive). These authors found that under the Kenward-Roger method, the average coverage estimate for the 95% confidence interval was the highest, .942, when autocorrelation was modeled versus not modeled. They also found that when autocorrelation was modeled, using Kenward-Roger method for estimating degrees of freedom, provided average coverage for the treatment effect that was close to the nominal level of .95.

A general conclusion based on the aforementioned studies is that the fixed effects are unbiased when using multi-level modeling with small sample sizes as long as the error structure and the degrees of freedom are correctly specified (Ferron, Farmer, & Owens, 2010). This conclusion was further investigated with a focus on the individual treatment effects and their confidence intervals when using one of three methods of estimating degrees of freedom: the Kenward-Roger, the Satterthwaite, or the Containment methods. Ferron, Farmer, and Owens (2010) concluded that traditional statistical methods, not accounting for the nested data structure, would tend to undercover with positive autocorrelation and the Kenward-Roger method would be expected to perform the best when there was a complex error structure. Furthermore, the authors summarized their article suggesting that researchers conducting multiple-baseline studies with multi-level modeling should use the Kenward-Roger method for estimating degrees of freedom (Ferron, Farmer, & Owens, 2010). Researchers are not only interested in analyzing single case designs across participants within a study, but there is also an increasing interest in analyzing data across single-case studies.

The Autocorrelation Debate

The repeated measures design is based on continuous observations over time for the same subject. This feature of single case research is a definite strength to this design given that it can allow a researcher to analyze a particular case in depth. However, this can also present challenges in terms of choosing an appropriate data analysis method. A variable whose future is predictable to some degree from its own values or from the passage of time possesses some form of statistical serial dependence. One way to describe autocorrelation is the extent to which the values of the observed behavior at time t (Y_t) are correlated with values at $t - i$, or Y $(_{t-i})$ (Matyas & Greenwood, 1996). This generally results in the data having characteristics different than taking an observation one or two times, such as in a between subjects design. This difference has implications on what assumptions (i.e. the assumption of independence) can be applied to the data and what statistical techniques can be used for analysis. The amount of dependence is typically characterized by the correlation between adjacent time points. This is referred to as autocorrelation, or serial dependency.

Whether or not single-case data can show serial dependence due to small sample sizes or how to best estimate the autoregressive parameters to ensure that they are unbiased due to the small *n* has been debated in the single case literature (Huitema & McKean, 1991). However, studies have shown that there is indeed some correlation beyond random chance in the repeated measures design for observations within a single subject (Kratochwill et al., 1974). The study's purpose was to demonstrate that the statistical independence assumption is entirely unwarranted in an N=1 (or small sample) design. The authors measured the correlation between time points within individuals compared to the correlation between heads or tails when flipping a coin. They found that the correlation was substantially higher between time points within the subject. This study supports the idea that what a person does at time t is not independent of what he or she had done at time t-1, t-2, t-3, etc. Many studies had continued to analyze the presence of autocorrelation in single case research (Busk & Marascuillo, 1988; Huitema & McKean, 1991;

Matyas & Greenwood, 1991). In another study, Busk & Marascuillo (1988) concluded that 40% of the baselines and 59% of the intervention phases had autocorrelation coefficients greater than .25. Huitema and McKean (1991) also confirmed the inappropriate choice for use in standardization used in Huitema (1985)'s original argument. However, the most important flaw with the original argument was that the conclusions may have been biased by the inclusion of many studies with very short time series (number of observations < 10), which may greatly underestimate the extent of autocorrelation (Matyas and Greenwood ,1991). Matyas and Greenwood (1991) also concluded that although the lag 1 autocorrelation may not be as large as originally hypothesized; it was clear that the general hypothesis of no autocorrelation cannot be sustained. Matyas and Greenwood (1996) conducted a review of the autocorrelation debate, synthesizing the different viewpoints regarding autocorrelation and single case designs. Based on their review, they concluded that neither visual nor statistical analyses could assume a simple "flat straight line plus random residual" model. Many of the aforementioned studies showed support for modeling the autocorrelation, or taking correlated errors into account when analyzing single case designs. Although there does seem to be agreement that yes, autocorrelation can exist among single-case data, there is still no agreement on the appropriate analysis technique to handle correlated errors in single-case data.

Recently, there has been an emergence of research dealing with autocorrelation in single case designs dealing with primary studies (Ferron et al., 2009; Ferron et al., 2010; Owens, 2011) and investigating intervention effects across studies, such as a meta-analysis (Baek & Ferron, 2013; Petit-Bois, Baek, & Ferron, 2013). These studies all have one thing in common: they all found that modeling autocorrelation tended to give more precise treatment effects versus not modeling the autocorrelation when autocorrelation was indeed present in the population. Given this conclusion, which analysis technique should be used to model autocorrelation when dealing with single-case data assuming that autocorrelation does exist?

Meta-Analysis and Singe-Case Designs

Treatment effectiveness, not only within studies but across studies as well, specifically which factors concerning the interventions are effective, has become a topic of great interest in terms of single case research (Beretvas & Chung, 2008; Owens, 2011; Van den Noortgate & Onghena, 2008). One method for addressing these concerns with treatment effectiveness across studies has been addressed through quantitative synthesis, or meta-analysis, of the research interventions to find which factors have been effective in math intervention research. Glass (1976) introduced the term meta-analysis as "the analysis of the results of statistical analyses for the purposes of drawing general conclusions" (p.3). Meta-analyses generally have three goals. 1) Meta-analytic studies strive to provide a point estimate of the average effect size, in order words, a quantitative summary. 2) Meta-analyses strive to provide confidence intervals in which the "true" population effect size is likely to be found. The confidence interval can then aid in the decision as to whether the effect size is significantly different from zero. 3) Meta-analytic techniques can help the researcher search for variables, or moderators, that could help explain the differences or variability among effect sizes, that is, given that there is a large amount or substantial variability among the effect sizes.

Single-case research has traditionally been left out of meta-analytic studies, due to the lack of agreement on the best method to meta-analyze single case data (Faith, Allison, Gorman, 1996; Van den Noortgate & Onghena, 2008). Meta-analysis of single case designs would not only allow for the understanding of generalizability, or the treatment effect across studies, but it also affords researchers the benefit of understanding how the treatment's effect relates to specific individuals within a particular study. Faith, Allison, Gorman (1996) offered additional reasons why it is essential to meta-analyze single case data. The authors posited that many interventions, particularly behavioral interventions have only been researched or studied in the single case context. Therefore, meta-analysis in single case research has to be done in order to know how effective these behavioral interventions have been across studies. Faith, Allison, Gorman (1996) further concluded that single-case research often leaves the reader wondering whether the results of a particular study could be applied to another individual outside of that context or study. Meta-analytic studies could then produce an average effect, with a confidence interval, which would then inform the reader what an expected effect size would be in other studies.

Analysis Methods

There are several methods used for conducting meta-analyses in single-case research. There are several meta-analyses that have been done involving single-case research designs; however, to date, there has not been a consensus on the best way to synthesize these data.

Summary Statistics. Beretvas and Chung (2008) conducted a narrative review of 25 single-case meta-analyses and found that most of the single case or single participant meta-analysts were using non-parametric approaches. They further concluded that meta-analysts were using the simplest indicators as effect size measures and non-parametric methods such as percentage of non-overlapping data (PND)(Scruggs & Mastropieri, 1998) and percent of all non-overlapping data (PAND) (Parker et al., 2007) for conducting meta-analysis involving single-case. Maggin et al. (2011) suggested that researchers should use both visual and statistical analysis when synthesizing across single-case research.

Regression-based Methods. Regression-type approaches are also commonly used to conduct meta-analyses because they offer sophisticated and flexible methods by fitting statistical models to the observed data (Faith, Allison, Gorman, 1996). Gorsuch (1983) concluded that trend analysis was the most important analysis in terms of minimizing your Type I and Type II errors. The major advantage to his method was that it modeled intervention effect over time. In other words, his method allowed meta-analysts to look at the intervention effect above and beyond the passing of time. The major shortcoming was that his method did not model change in slopes, only change in levels. However, there are interventions where the researcher does not only expect there to be a change in level, but also for there to be a change in slope or trend. Several researchers have then proposed statistical models that would allow the researcher to determine treatment effectiveness when a trend is present in the data (Allison & Gorman, 1993; Kromrey & Foster-Johnson, 1996).

These effect sizes are much more complex. This is due largely in part by the fact that trend must be taken into account before analyzing intervention effectiveness (Kromery & Foster-Johnson, 1996). The authors explicate a method for computing effect sizes, for either shift in levels or change in slopes, using a regression based approach when trend is present in the data. Kromery and Foster-Johnson (1996) demonstrated that the effect size can be calculated using the change of R^2 for the two models--the second model taking into account the trend while the initial model does not take trend into account. In addition to these well-known approaches, some researchers have turned to the use of multi-level modeling as an additional statistical tool for synthesizing single case data across studies (Owens, 2011; Ugille et al., 2012; Van den Noortgate & Onghena, 2008).

Multi-level Modeling. In addition to the single-variable effect size indicators and the previously mentioned regression type indicators that were being used to meta-analyze single-case designs, Van den Noortgate and Onghena (2008) proposed that if raw data from several single case designs are used in a meta-analysis, scores can be varied at each of the three levels: over occasions (level-one), across participants from the same study (level-two), and across studies included in the meta-analysis (level-three). Equation 13 below describes the variation within participants that occurs when treatment conditions are compared with a baseline condition (Level 1) allowing trends for both the baseline and treatment phases. At the second level, the variation over participants is shown using four regression equations (Equations 14-17). Finally, the last set of equations describes the variation across the studies (Equations 18-21) that are included in this meta-analysis.

Equations 13-21 below represent the three-level model used to represent the fixed effects and the variance components at each of the three levels. The variable phase is a dichotomous variable representing the baseline phase ($phase_{ijk}$=0) and the treatment phase ($phase_{ijk}$ =1).

Level 1 Equation:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}{*}phase_{ijk} + \beta_{2jk}{*}time_{ijk} + \beta_{3jk}{*}phase_{ijk}{*}time'_{ijk} + e_{ijk}, \quad e_{ijk} \sim N(0, \Sigma_e) \quad (13)$$

Level 2 Equations:

$$\beta_{0jk} = \theta_{00k} + u_{0jk} \quad (14)$$

$$\beta_{1jk} = \theta_{10k} + u_{1jk} \quad (15)$$

$$\beta_{2jk} = \theta_{20k} + u_{2jk} \quad (16)$$

$$\beta_{3jk} = \theta_{30k} + u_{3jk} \quad (17)$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim N(0, \Sigma_u)$$

Level 3 Equations:

$$\theta_{00k} = \square_{000} + v_{00k} \qquad (18)$$

$$\theta_{10k} = \square_{100} + v_{10k} \qquad \begin{bmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v_{30k} \end{bmatrix} \sim N(0, \Sigma_v) \qquad (19)$$

$$\theta_{20k} = \square_{200} + v_{20k} \qquad (20)$$

$$\theta_{30k} = \square_{300} + v_{30k} \qquad (21)$$

Combined Model:

$$Y_{ijk} = \square_{000} + \square_{100}\text{*phase} + \square_{200}\text{*time} + \square_{300}\text{*phase*time'} + v_{00k} + v_{10k} + v_{20k} + v_{30k} + u_{00k} + u_{10k} + u_{20k} + u_{30k} + e_{ijk} \qquad (22)$$

In Equation 13, the value of the dependent variable on measurement occasion *i* for subject *j* of study *k* (Y_{ijk}) is regressed on a dummy variable phase, that equals one if measurement occasion *i* occurs in the treatment phase, otherwise it is equal to 0 in the baseline phase. The score on the dependent variable on measurement occasion *i* for subject *j* of study *k* is Y_{ijk}. The *phase* variable is a dummy-coded indicator reflecting whether the observation is in the baseline phase ($phase_{ijk} = 0$) or the intervention phase ($phase_{ijk} = 1$). The coefficient for the *time* variable represents the slope during the baseline phase, $time_{ijk}$,and the coefficient for the interaction term, $phase_{ijk}*time'_{ijk}$, reflects the difference between the baseline and intervention phases' slopes. The *time* variable can be centered in a variety of ways which may be helpful for interpreting the model parameters (Baek et al., 2013; Van den Noortgate, & Ferron, 2013).

Specifically, for this study, the *time* variable is uncentered for the baseline phase's slope such that for the first measurement occasion in the baseline phase, $time_{ijk} = 0$. However, the *time′* variable in the interaction term, $phase_{ijk}*time'_{ijk}$ is coded such that $time'_{ijk} = 0$ for the first observation in the intervention phase. Therefore, the expected score during the baseline phase if extended one observation into treatment is equal to β_{0jk} ; and the expected score of the treatment phase at this same point in time is β_{1jk} higher. As a result, β_{1jk} refers to the immediate level change associated with an intervention (Huitema & McKean, 2000). β_{2jk} represents the linear trend during the baseline phase, and the linear trend for the treatment phase was β_{3jk} higher. In other words, β_{3jk} indicates the effect of the intervention on the trend that is the difference between the baseline phase and treatment phase slopes. Both β_{1jk} and β_{3jk} are needed to fully describe the treatment effect. At the second level, Equations 14-17 describe the new regression equations for the variation over subjects. Equation 14 describes that the baseline performance for subject j from study k equals an overall baseline for study k plus some random deviation. Similarly, Equations 15-17 indicate the variation of the treatment effect, linear trend in baseline, and the effect of the intervention on linear trend, respectively, over subjects from the same study. The next set of equations can be thought of similarly as the second level equations.

At the third level, the variation across subjects is modeled using Equations 18-21. Equation 18 represents the baseline mean for study *k* as the overall baseline across all of the studies plus some random deviation. The same is modeled by Equations 19-21 for the variation of the treatment effect, linear trend in baseline, and the effect of the intervention on linear trend, respectively, across studies. It should be noted that errors on levels 2 and 3 are typically assumed to be normally distributed and have a mean of 0 and a variance of 1.0. There was no covariance in the errors between levels and between errors at level-two and level-three. The within-person error is modeled by e_{ijk} (σ^2_e represent the variance of e_{ijk}). Equation 22 simply represents the combined model once all coefficients have been algebraically substituted.

There are lingering concerns on the use of multi-level modeling, which is based on large sample theory, and its appropriateness to single case data. One may expect that interval estimates of the average treatment effects would be unbiased under smaller sample sizes;

however, the same is not expected of the variance components (Raudenbush & Bryk, 2002). Owens (2011) conducted a Monte Carlo study to examine the appropriateness of the three-level model to meta-analyze raw data from single case studies. The author found that the fixed effects estimated in a three-level model tended to be reliable and reasonably unbiased with small sample size and when using the Kenward-Roger estimation for degrees of freedom. In short, the author found that the 95% confidence intervals width for the fixed effects approached .95 as level-three sample size increased. This indicated that, whenever possible, meta-analysts should increase the number of primary studies included in their meta-analysis. The simulation study also found that variance components tended to be less stable and more biased. Specifically, Owens (2011) found the level-three variance components tended to be underestimated, while the level-two variance components tended to be overestimated. Furthermore, as the variance in error terms shifted from most of the variance being in level-two to most of the variance being in level-three, the variance components for level-three tended to be underestimated and more biased. Conversely, as the variance in error terms shifted from most of the variance being at level-one to most of the variance at level-two, the variance components for level-two tended to be overestimated and more biased. Owens (2011) also found that the within person residual variance became more biased when using the three-level model as autocorrelation increased. The author proposed that this finding was not a surprise, given the stronger the relationship between errors within the person, the greater the difficulty in obtaining unbiased estimates for the residual variance.

Ugille et al. (2012) conducted an extensive simulation study which investigated the performance of the multi-level approach for standardized (the unstandardized regression coefficients divided by the residual within-phase standard deviation) and unstandardized effect sizes for single case studies. The authors simulated various conditions for the three-level model. They found that the multi-level approach worked well when unstandardized effect sizes were used. The approach was also optimal for standardized effect sizes for certain conditions: when there were more than thirty studies, when there were many level-one units, i.e. observations for each participant and when the studies are rather homogeneous, and when there was a small amount of between-study variance.

Furthermore, applied work has been done examining the three-level model to meta-analyze single case data. Petit-Bois, Baek, and Ferron (2012) investigated the model by analyzing the degree to which parameter estimates are sensitive to various methodological decisions, specifically regarding the specification of the growth trajectories using raw data collected from primary studies. Three distinct models involving different specifications of growth trajectories (no growth within a phase, constant linear growth, or nonlinear growth) was analyzed to understand the impact of this methodological decision on the parameter estimates. The study found that the model did support the visual analysis graphs in selecting the best model (the model that specified linear growth in terms of mathematics achievement over time). The authors suggested future work could be done to analyze the appropriateness of the fit indices given the small samples associated with single case research (Petit-Bois, Baek, and Ferron, 2012). In another study conducted by Baek, Petit-Bois, and Ferron (2012), the three-level model was evaluated looking at the consequences of error structure specification on the results of a meta-analysis of single-case data involving reading fluency. More specifically, the authors analyzed four different models: the first model assumed no autocorrelation; the second model the autocorrelation is assumed to be constant both within and across studies; the third model the autocorrelation is assumed to be the same across participants within a study, but allowed to vary across studies; and the last model allowed for varying autocorrelation across participants which leads autocorrelation to vary both within and across studies. The results indicated that the last two models did not converge; therefore, the remaining results are based on just the first two models. The fixed effects for the two models did not statistically differ. However, the fit indices supported the more complex level-one error structure (Baek, Petit-Bois, & Ferron, 2012). These applied works did use the fit indices as one way to identify the best model to approximate the data, however, the appropriateness of the use of fit indices have not been empirically studied in terms of single case research.

Based on the research done thus far investigating Van den Noortgate's 2008 three-level meta-analytic model, this study sought find the appropriateness of the use of fit indices to select the correct level-one error structure when synthesizing raw data across single-case studies. Furthermore, this study looked at the consequences of misspecifying the level-one error structure when using the three-level model to meta-analyze single case data. This work has not yet been done in single-case research, or using small sample sizes, therefore literature from the broader repeated measures or longitudinal designs was used to further inform the conditions for this study.

Level-one Error Specification

There are several ways that autocorrelation can be modeled, however, this study focused on a few which are discussed in the following section.

Error Structure Options

The simplest approach to modeling the level-one error structure is to assume the errors are independent, $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ (Raudenbush & Bryk, 2002). Another simple alternative would be to specify an unstructured covariance matrix. The great appeal for an unstructured error covariance is that it places no restrictions on the structure of **Σ**. An unstructured covariance structure is commonly used in longitudinal data analysis, where there are generally a large number of participants and a substantial number of observations per participant. In most analyses, a more parsimonious structure is desirable. For example, in an exploratory analysis, it is sensible to begin with the unstructured error covariance model because it has the smallest deviance. This is due to the large number of parameters that are required to be estimated. The large value for the AIC and BIC over the model that assumed independent errors demonstrates the "wasting" of considerable degrees of freedom in choosing an unstructured form of **Σ** (Singer & Willet, 2003). However, convergence issues may arise with the use of the unstructured matrix in single case research where there are typically a small number of participants with a large number of observations for each participant.

There are several other types of error matrices used in the repeated measures literature. One type of level-one error structure would be to use a first order autoregressive model, AR (1). Conceptually, the lag 1 autocorrelation represents the degree to which the current observation, at time t, can be predicted by the observation before, at time t-1. This can be calculated by taking the correlation between the second and first observation and so on, throughout the series. By doing this a researcher can determine the relatedness of the current observation with the past observations. Many researchers are drawn to the first-order autoregressive model because its "banded diagonal" shape seems appropriate or realistic for growth processes. When errors are characterized by AR(1), the elements in the main diagonal of **Σ** have equal variances (homoscedastic, with variance σ^2). Additionally, the pairs of errors have identical covariances in bands parallel to the leading diagonal. The covariances are the product of the σ^2 and an autocorrelation coefficient, ρ, whose value is always less than or equal to 1. Due to the fact that the errors are always fractional, then the error variances decline as you move away from the leading diagonal (Singer & Willet, 2003).

There are also more complicated models, such as the first order autoregressive moving average model, ARMA (1,1). This alternative has some characteristics of the autoregressive structure, in that it has bands of identical covariances aligned parallel to the main diagonal. The ARMA (1,1) allows more flexibility than the AR(1) structure (Singer & Willet, 2003,). In summary, the list of parameters for each model can be described as ID contains a single parameter (σ^2) on the main diagonal of an identity matrix, whereas AR(1) contains two parameters (σ^2 and the autocorrelation coefficient, rho, ρ), and ARMA (1,1) contains not only the same two parameters as in AR(1), σ^2 and ρ**,** it also has a moving average coefficient, gamma (γ). These structures are considered nested because one structure can easily be another structure by constraining one or more parameters. More specifically, AR(1) can be reduced to ID if ρ is constrained to be equal to 0. Moreover, ARMA (1,1) can be constrained to ID if both γ and ρ are set to equal 0 (Kwok, West, & Green, 2007). This is illustrated by the three level-one error matrices in Figure 3 below. This example is just for simple illustration in Figure 3 below of the conceptual relationship between these nested covariance structures.

$$\mathbf{\Sigma} = \sigma^2\mathbf{I} \qquad \left|\begin{matrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{matrix}\right|$$

$$\mathbf{\Sigma} = \text{AR}(1) \qquad \sigma^2\left|\begin{matrix} 1 & P & \rho^2 & \rho^3 \\ P & 1 & \rho & \rho^2 \\ \rho^2 & P & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{matrix}\right|$$

$$\mathbf{\Sigma} = \text{ARMA}(1) \qquad \sigma^2\left|\begin{matrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \Gamma\rho \\ \gamma\rho & \gamma & 1 & \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{matrix}\right|$$

Figure 3. Three level-one error structures. This figure illustrates the 3 different error specifications for the level-one error structure. Furthermore, it illustrates the correlation across four time points.

Another way to illustrate these structures is by looking at the underlying models which are often presented in the time series literature. Tabachnick & Fidell (2008) describes autoregressive components as the memory of the process preceding observations. The value of phi (ρ) is 0 when there is no relationship between adjacent observations. Furthermore, they described the relationship using the following mathematical model in Equation 23:

$$Yt = \rho Y_{t-1} + e_t \qquad (23)$$

The moving average components represent the memory of the process for preceding random errors (Tabachnick & Fidell, 2008). The authors went on to describe the mixed model which contained both an autoregressive and moving average components so both types are required for this model. This mixed model is represented below in Equation 24 where ρ is the autoregressive component and □ illustrates the moving average component:

$$Y_t = \rho Y_{t-1} - \square e_{t-1} + e_t \qquad (24)$$

Obtaining the correct within-subject covariance structure has been a huge discussion and studied extensively in growth curve modeling (GC modeling) or longitudinal data analysis (Kesselman, Littell, & Sivo, 2000). Growth curve modeling or longitudinal data's defining characteristic is that individuals are measured repeatedly over time enabling direct study of change (Diggle, Heagerty, Liang, & Zeger, 2002). Should one use a correlated error structure? Many researchers who use uncorrelated error structures commonly assume that the $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ (Bryk & Raudenbush, 2002). This commonly used approach should lead the researcher to question whether $\mathbf{\Sigma}$ has been misspecified (Kesselman, Algina, Kowalchuk, & Wolfinger, 1999; Kwok, West, Green, 2007). Given this question and the research which has demonstrated that autocorrelation more than likely exists within a repeated measures framework, choosing to model the autocorrelation appears most appropriate. After deciding to model autocorrelation, a question that commonly arises with both areas (whether it is longitudinal or single case designs) is whether or not the model has the correct level-one error structure specification?

Selecting the Correct Level-One Error Structure

There are two commonly used approaches to select an appropriate level-one covariance structure. One method is to use fit indices to appropriately select the best model to approximate the data. The other option is to specify the level-one error structure a priori with either an uncorrelated or correlated error structure. Both of these methods have been studied significantly in terms of longitudinal data or growth curve modeling.

Fit criteria. The first method is to allow the fit information to select the appropriate error structure. Several researchers have utilized this method of selecting a structure for **Σ** by examining multiple structures and using log likelihood tests or information criteria to select an error structure (Ferron, Dailey, & Yi, 2002; Gomez, 2005; Kesselman et al., 1999). Commonly used fit indices include deviance statistic, AIC, or BIC (Ferron, Dailey, & Yi, 2002; Singer & Willet, 2003).

There are several advantages to using index comparison approaches, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). One advantage is that it allows the comparison of non-nested models. Another benefit to using to index comparison is that the indices quantify the degree to which the model represents an improvement over the competing models (McCoach & Black, 2008). Additionally, Liu, Rovine, and Molenar (2012) also suggested that the most general approach to compare models regardless of the type of misspecification is to look at the AIC and BIC. In their study, the authors hypothesized that there could be a difference between the structural equation modeling (SEM) approach and the hierarchical linear modeling (HLM) approach. They theorized that the difference could stem from the method by which each approach estimated the parameters. For example, the SEM approach can separately model the means and the covariance parts and evaluate the fit of the covariance model alone; meanwhile, the HLM approach simultaneously estimates the means model and the covariance model (Liu, Rovine, & Molenar, 2012). However, they found that the fit indices performed well with both small and large sample sizes.

The formulas for the two indices appear similar, but the indices are philosophically different. The Bayesian approach, or BIC, treats every model as the possible "true" model, and estimates the likelihood of the model being the correct model. For the AIC, this index uses the prediction of future data as the key criterion (McCoach & Black, 2008). Although some researchers may prefer one index over the other, McCoach & Black (2008) suggest the use of both the AIC and BIC in combination with chi-square difference tests for nested models. The formulas for the three fit indices are shown below:

$$\text{AIC} = \text{D} + 2p \tag{25}$$

$$\text{AICc} = \text{D} + 2\text{p} * (n \,/\, (\, n - p - 1)) \tag{26}$$

$$\text{BIC} = \text{D} + \ln\,(n)*p \tag{27}$$

For all of the indices, D represents the deviance (-2LL). However, for the AIC, p = the number of parameters estimated in the model. Since the AIC's penalty term is 2p the deviance must decrease by more than 2 per additional parameter, in order to favor the more complex model. For the BIC, n represents the sample size and p is the number of parameters estimated in the model. In multi-level models, it is not clear which sample size should be used. SAS PROC MIXED uses the number of independent sampling units as the sample size. At small sample sizes, the BIC will favor more parsimonious models than the AIC and the chi-square difference tests.

Furthermore, Raftery (1995) has suggested guidelines for interpreting changes in BIC. After subtracting the BICs for the two competing models, a difference of 0-2 suggests weak evidence favoring model 2 over model 1; differences of 2-6 provides positive evidence for model 2 over model 1; BIC differences of 6-10 provided strong evidence for the model 2 over model 1; and lastly, differences greater than 10 provided very strong evidence for model 2 over model 1. For the AICc, a finite-sample corrected version of the AIC, where n represents sample size. Therefore, AICc gives a greater penalty than the AIC for extra parameters.

Ferron, Dailey, and Yi (2002) wanted to analyze the sensitivity of model selection criteria to the misspecification of the level-one error structure. The effects of the misspecification were then examined for estimates of variance parameters, estimates of the fixed effects, and the tests of fixed effects. They found that the fixed effects were not biased. This finding aligned with past simulation work. They also found that the AIC correctly identified only 47% of the time. Their results varied greatly as a function of sample size (larger sample sizes gave more precise estimates). There was also a notable interaction between the series length and sample size (sample size matter more when there were shorter series length). The implication of this study was that the fit indices do not properly identify the correct error structure. This study also demonstrated that if the error structure was not modeled correctly, then this would lead to even more bias in the variance components.

Another study that investigated fit indices was Kesselman, Algina, Kowalchuk, & Wolfinger (1999). This study sought to compare two different types of fit indices for selecting covariance structures when looking at repeated measures designs. They found that neither fit index, the Akaike (1974) or the Schwartz (1978), uniformly chose the correct covariance structure. This study indicated that although there is a need to model the correct covariance structure, due to powerful tests of the fixed effect parameters, using fit indices is not a reliable method for choosing the correct error structure. These findings are very similar to Ferron, Dailey, and Yi (2002); both studies suggested there is a lack of reliability in using fit indices to correctly identify the correct error structure in repeated measures designs.

Selecting error structure a priori. The second method to select a covariance structure for $\mathbf{\Sigma}$ a priori has also been studied substantially in the longitudinal area. Sivo, Fan, and Witta (2005) investigated the degree to which autocorrelation in its various forms biases the estimates obtained in latent GC modeling. This study had two intended purposes: 1) to introduce how growth curve models and MA and ARMA models may be integrated; and 2) to investigate the degree to which autocorrelation in its various forms (AR, MA, and ARMA) biases the estimates in GC modeling. They found that unmodeled autocorrelation could lead to biased results. Their suggestion was to always model autocorrelation as an option to improve model fit when applying GC modeling with at least 4 time points in longitudinal research. The authors further suggested more work was necessary that focused on fewer occasions to determine whether the conclusions of the study hold. Another study, Kwok, West, and Green (2007) looked at the effects of different forms of misspecification (underspecification, general misspecification, and overspecification) of the within-subject residuals for longitudinal models. They found that underspecification and general misspecification of the level-one error matrices were more likely to result in overestimation of the standard errors of the growth parameters, which resulted in lower statistical power as compared to the correct specification. As a result, the authors concluded that overestimation of the matrices were more likely to result in slightly smaller standard errors of the growth parameters which led to a possible gain in statistical power. These findings led the authors to suggest that it is best to adopt a slightly overspecified, such as AR(1), error structure if researchers are not sure about the correct error structure for their data.

In a later study, authors examined the performance of a two-level model when autocorrelation moving average is present, but the data were misspecified and modeled as either σ^2 **I,** first-order autocorrelation, or unstructured covariance matrix (Murphy & Pituch, 2009). Some key findings of this study were that the fixed effects were unbiased as was found in previous research (Ferron, Dailey, & Yi, 2002; Raudenbush & Bryk, 2002). The authors also found that overspecifying the level-one error structure with type = UN was a viable option with sufficient sample size. Finally, the authors concluded that the variance components were biased regardless of correct specification. This research again suggested that when serial dependence or autocorrelation is present, the fit criteria will not always correctly identify the covariance structure. Due to this lack of reliability of the fit indices, the author chose to always overspecify the error structure using type = UN. However, Singer and Willet (2003) explicated that overspecifying the covariance matrix using the unstructured matrix was not typically ideal. The authors suggested the desire for the more parsimonious model. The authors explicated that their study was an example of fitting various types of matrices, including compound symmetry, first order autoregressive, and a toeplitz, and then using the fit indices to determine the correct error structure. They found that the toeplitz error structure most appropriately fit their data based on the results of the fit indices.

Sensitivity analysis. Additionally, researchers can perform some sort of sensitivity analysis, where they can fit several level-one covariance structures and see how these different models affect the precision of the parameter estimates. Based on the consequences to the estimates, then a research can choose the most accurate level-one error structure. Faith, Allison, and Gorman (1996) recommended that meta-analysts may try to look at the effect sizes of one method, then repeating analyses using another error structure specification as a form of sensitivity analysis.

A general conclusion for the literature from the broader repeated measures area that can be drawn was that researchers do not typically know what type of error structure would best approximate their data. Therefore, researchers typically have three options: 1) to use fit information criteria to select the correct level-one error structure or 2) to choose their error structure a priori. 3) To perform some form of sensitivity analysis.

Chapter Summary

Single-case designs are used extensively to determine treatment effectiveness or how a treatment may affect a single subject or multiple participants within a study. There are several commonly used designs in single case research. The most popular design is the multiple-baseline (Shadish & Sullivan, 2008). This design is powerful due to its ability to reduce internal validity threats, or the possibility that anything other than the treatment or intervention could be causing the participants to change the observed behavior or outcome at the time the intervention is introduced (Barlow, Nock, & Hersen, 2009; Kazdin, 2011).

One important feature of single-case interventions is the repeated measurement of a case or participant across multiple time points, to assess the treatment's effect on one specific case or participant (Kazdin, 2011). These repeated observations within one participant have led to a great debate about how correlated one observation is to the next observation and so on and so forth. Autocorrelation can affect how the researcher interprets their intervention effect, no matter the type of analysis that is used (Barlow, Nock, & Hersen, 2009; Kazdin, 2011). Visual analysis is the most commonly used technique applied to single case designs; however, this analysis has limitations in the presence of correlated level-one error structures (Brossart et al., 2006; Huitema, 1986; Kazdin, 2011; Maytas & Greenwood,1990). Several researchers have suggested the need for visual analysis to be supplemented by some statistical technique (Barlow, Nock, & Hersen, 2009; Kazdin, 2011). Nevertheless, some of the commonly used statistical techniques also show concerns in the presence of correlated errors.

The multi-level model allows flexibility in modeling autocorrelation; moreover, the model allows for different level-one error specifications. Single-case researchers are occasionally interested in more than synthesizing data across participants within a study; they are also interested in synthesizing across studies. There are several meta-analyses that have been done involving single-case research designs; nonetheless, there still has not been a consensus on the best way to synthesize these data. Van den Noorgate (2008) proposed a three-level model that can be used to synthesize raw data across single-case studies. There are several ways to specify the three-level model, and one of those decisions is whether or not to model autocorrelation. If choosing to model autocorrelation, how do we know that the level-one error structure has been correctly specified in the model? This work has not yet been studied in the single-case literature, however, literature from the broader repeated measures or longitudinal area was utilized to inform this study.

Obtaining the correct within-subject covariance structure has been a huge discussion and studied extensively in growth curve modeling (GC modeling) or longitudinal data analysis

(Kesselman, Littell, & Sivo, 2000). There are two commonly used approaches to the selection of an appropriate level-one covariance structure when using multi-level models. One method is to use fit indices to appropriately select the best model to approximate the data (Ferron, Dailey & Yi, 2002; Gomez, 2005; Kesselman et al., 1999). The other option is to specify the level-one error structure a priori with either an uncorrelated or correlated error structure (Kwok, West, & Green, 2007; Murphy & Pituch, 2009).

The study examined at the appropriateness of the use of fit indices to correctly identify the level-one error structure. More specifically, the study analyzed the percentage of times that each fit index appropriately guides the researcher to the correct level-one error structure. Secondly, the study looked at the consequences of misspecifying the level-one error structure when synthesizing data across single-case studies when utilizing the three-level model.

CHAPTER THREE: METHOD

This chapter outlines the proposed methods for this study, including the purpose, research questions, sample, design, and analysis.

Purpose

There have been a number of studies that have analyzed fit indices and properly identifying the correct level-one covariance structures in terms of the general longitudinal data, or growth curve models. A thorough literature review has uncovered no research to date looking at fit indices in terms of single-case research using multi-level models. Moreover, no research was uncovered regarding the consequences of different forms of specification of the level-one error structure when using a three-level meta-analytic single-case model.

The purpose of the study was two-fold: 1) to determine the extent to which the various fit indices (post hoc selection) can correctly identify the level-one covariance structure, and 2) to investigate the effect of various forms of misspecification of the level-one error structure when using a three-level meta-analytic single-case model.

Research Questions

1. To what extent do fit indices (AIC, adjusted AIC, BIC, log likelihood ratio test) correctly identify level-one covariance structure when using a three-level meta-analytic single-case model?
2. To what extent are the **fixed effect** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary

study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?

3. To what extent are **confidence interval width and coverage for the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?
4. To what extent are the **Type I error and power for the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification)?
5. To what extent are the **variance component** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, and covariance structures, level of the treatment effect), and analysis factors (form of specification)?
6. To what extent are **confidence interval width and coverage for the variance components** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of the treatment effect), and analysis factors (form of specification).

Design

The study used a 2x2x2x2x2x5x7 factorial design. Seven experimental variables were manipulated in this study: 1) the number of primary studies per meta-analysis (10 and 30); 2) the number of participants per primary study (4 and 8); 3) the series length or number of observations per participant (10 and 20); 4) variances of the error terms (most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at higher levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]); 5) the level for the fixed effects (0 or [2 for the shift in level and .2 for the shift in slope]); 6) the level of autocorrelation and the moving average parameter, respectively: [(0,0), (.2, 0), (.4,0), (.2, .2), (.4, .4)]; and 7) the form of model specification [i.e. ID, AR(1), ARMA (1,1)], and error structure selected by AIC, AICC, BIC, and the LRT. The next section will provide a thorough description of the conditions that were simulated in this study.

Conditions Simulated

Number of primary studies per meta-analysis. These values were chosen based on several initial studies that have been conducted in this area. First, a review that was conducted by Farmer, Owens, Ferron, and Allsopp (2010) on 39 single-case meta-analyses in social science between the years of 1999 and 2009 found that the number of primary studies included in the meta-analyses ranged from 3 to 117. Additionally, Farmer et al. found that 60% of the meta-analyses included less than 30 primary studies. Owens and Ferron (2011) conducted an initial study that looked at the three-level analytic model synthesizing raw data from single-case studies and used three levels for the number of primary studies per meta-analysis: 10, 30, and 80. The authors found that the confidence interval around the fixed effects approached the nominal level of .95 as the number of studies included in the meta-analysis increased. In another study, the number of primary studies used was 10 and 30 (Ugille et al., 2012). The authors found similar results that the confidence interval became more precise as the level-three sample size increased.

Initial applied work was also done to further inform the conditions used in this study. Petit-Bois (2012) conducted a meta-analysis of mathematics interventions and found only 10 studies that used comparable dependent variables, mathematical problem solving, and met the inclusion criteria of the study. Another applied meta-analysis using the raw data from single-case studies included the dependent variable, reading fluency, and in this study, Baek, Petit-Bois, and Ferron (2012) found 20 studies that appropriately met the inclusion criteria. The conditions that were used in this study were based on all of the aforementioned work that has been done thus far in this area. The number of primary studies in each meta-analysis had two levels: 10 and 30.

Number of participants per primary study. The survey of single-case meta-analyses found that the majority of single-case studies involved less than or equal to 7 participants (Farmer, Owens, Ferron, & Allsopp, 2010). Kazdin (2011) recommends that in order to see a treatment effect "clearly" that a minimum of three baselines is necessary. However, the author goes on to explain that the more baselines there are, the "clearer" the intervention effect is. Specifically, intervention effects are more evident across several (8 or 9) persons or situations (Kazdin, 2011). An overview of the simulation work that has been conducted revealed that in two other studies, Owens and Ferron (2011) used two levels for the modal number of participants: 4 or 8; and the other study used 4 or 7 seven participants for their Monte Carlo study (Ugille et al., 2012).

A further examination into the applied work that has been conducted in this area exposed that the average number of participants for each of the primary studies was 3.25 (Petit-Bois,

2012) and 4.60 (Baek, Petit-Bois, & Ferron , 2012). In a survey of the characteristics of single case designs, Shadish and Sullivan (2011) found that the sample sizes ranged from 1 to 13, with an average of 3.64. Based on these findings and the previous work that has been done in terms of Monte Carlo studies, there were two levels applied for number of participants: 4 and 8.

Series length per primary study. Shadish and Sullivan (2011) found that of the 809 studies that were reviewed over 90% had 49 or fewer observations. Moreover, previous simulation work using the two-level model for analyzing single case data used series length of 10, 20, and 30 (Ferron et al., 2009; Ferron, Farmer , & Owens, 2010). In the Monte Carlo studies, the series length had three levels for one of the studies: 10, 20, and 30 (Owens & Ferron, 2011); another study that was recently conducted included the series length with two levels: 10 and 30 (Ugille Moeyaert, Beretvas, Ferron, & Van den Noortgate, 2012). The authors in the latter study, in 2012, Ugille et al. found that the most bias conditions in their data set were the conditions that contained the shortest series length, level-one sample size of 10. In the area of applied work, Petit-Bois (2012) included studies with an average series length of 12.55. These prior studies were used to determine the series length for this study. The series length for this study included series lengths of 10 and 20.

Baseline stability (lacking new trends or high variability) is most important when introducing the second baseline and any other consecutive baselines to intervention. The focus should not be on the number of days, but rather on the clarity of the pattern in determining the decision (Kazdin, 2011). The time at which the intervention was introduced staggers across participants within studies, creating the multiple baselines for the study. This time will depend on the combination of the number of participants and data points for each. These combinations are reflected in the table below. For example, when there are four participants the number of measurement occasions for each participant equals ten. For the first participant, the treatment began on the fourth observation; for the second participant on the fifth observation; for the third participant on the seventh observation; and for the last participant, the treatment began on the eighth observation. The treatment will last until the tenth measurement occasion for all four participants. For the studies that include eight participants, the eight participants were paired to form 4 groups. These 4 groups of dyads will then be on the same intervention schedule as the studies that only have 4 total participants. For example, the first dyad, began treatment on the fourth observation (4); for the second dyad on the fifth measurement occasion (5), the third dyad on the seventh observation (7), and for the last pair, the treatment began on the eighth observation (8). This is represented as 4-5-7-8 in the table below. The treatment will last until the tenth measurement occasion for all four pairs of dyads.

Table 2 below illustrates the multiple baselines for all four combinations of the number of participants and total number of observations or measurement occasions.

Table 2
The Combination of Treatment Introductions for the Various Numbers of Participants and Observations

		Number of Participants for each study	
		4	8 (4 pairs)
Total Number of Observations	10	4-5-7-8	4-5-7-8
	20	6-9-12-15	6-9-12-15

Variances of the error terms. The variances could be separated into two distinct groups: either having most of the variance at level-one, variance within participants (Ferron et al., 2009; Van den Noortgate, 2008) or having most of the variance at the higher levels, the variance among the participants and the variance among studies included in the meta-analysis (Van den Noortgate, 2008). Based on these findings, elements of the within-study (across participants) variance matrix, Σ_u, were manipulated to represent both scenarios. For simplicity, the covariances or off-diagonals were set to 0 (there was no covariances) for the within-study (across participants) variance matrix. Therefore, Σ_u, is a diagonal matrix, $\Sigma_u = diag(\sigma^2_{u_0}, \sigma^2_{u_1}, \sigma^2_{u_2}, \sigma^2_{u_3})$. If the within-person (level-one) variance is set to 1.0, setting the four diagonal elements of Σ_u to values of 2, 0.2, 2, 0.2 (for the variances in the baseline's intercept, baseline's slope, shift in level, and difference in baseline and treatment's slope, respectively) represents a relatively large amount of within-study (level-two) variability. Conversely, setting the four diagonal elements of Σ_u to values of 0.5, 0.05, 0.5, 0.05 (for the variances in the baseline's intercept, baseline's slope, shift in level, and difference in baseline and treatment's slope residuals, respectively) represents a relatively large amount of within-person (level-one) variability.

The same idea was applied to the level-three variance matrix, the variance across, or between-studies. The elements of this variance matrix were $\Sigma_v = \text{diag}(\sigma^2_{v_0}, \sigma^2_{v_1}, \sigma^2_{v_2}, \sigma^2_{v_3})$ equal to 2, 0.2, 2, 0.2 to represent a relatively large amount of between-study variability (level-three) and 0.5, 0.05, 0.5, 0.05 to demonstrate a relatively small amount of between-study variability. Owens (2011) found that the level-two variance components tended to overcover when most of the variance was at level-one, however, the level-three variance components tended to undercover when most of the variance was at level-one. The study applied similar conditions for the error variances. Therefore, the study had two conditions, most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at the higher levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]. Note that the variance of the error terms for the residual variance was held constant at 1.0 across conditions. Thus, the conditions that have most of the variance shifted to the upper levels have more total variance.

Levels for the fixed effects. The level for the treatment effect was typically fixed in prior research, in which the aim in those studies was not to look at the power estimates for the study. Those prior works (Ferron et al.,2009; Owens, 2011) were focused on looking at the bias in the point estimates and interval coverage for the fixed effects and the variance components (Baek & Ferron, 2013; Owens, 2011). However, for this study, the power estimates and the type I errors were outcomes of interest, therefore two levels for the fixed effects were selected. The levels for the fixed effects were either no effect, i.e. 0, for both the shift in level and the shift in slope. Alternatively, the levels for the fixed effects were 2 and 0.2 for the shift in level and shift in slopes, respectively.

Level-one error structures. The data were generated using five different level-one error structures. First, the simplest error structure was used; this structure will assume no autocorrelation, $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ (Raudenbush & Bryk, 2002). The second error structure that was generated assumed a first-order autoregressive structure, AR(1). In past simulation work that examined autocorrelation in terms of the multi-level model, specifically the two-level model in this case, the levels of autocorrelation were 0, 0.1, 0.2, 0.3, 0.4 (Ferron et al, 2009; Ferron, Farmer, & Owens, 2010). These values covered the range that is typically found in behavioral or educational research (Huitema, 1985; Matyas & Greenwood, 1996). Likewise, there have been additional Monte Carlo studies that have investigated the appropriateness of the three-level model with synthesizing raw data across single-case studies, the levels of autocorrelation used were 0, 0.2, and 0.4 (Owens, 2011). The author found that within-person residual variance became more biased as the level of autocorrelation increased. For this study, there was five levels for the autocorrelation and moving average parameters, respectively: the ID model: [0, 0], autoregressive model: [.2, 0], autoregressive model: [.4, 0], moving average autoregressive model: [.2, .2], and moving average autoregressive model: [.4, .4]. These values are commonly used in educational research.

Form of model specification. The study evaluated the consequences of multiple approaches to level-one error structure specification, including post hoc specification based on a range of fit indices and *a priori* specification of relatively simple to relatively complex covariance structures. These methods were chosen to parallel the options used by analysts in practice and in past research in broader longitudinal research. The use of fit indices or post hoc model selection has been extensively examined in the broader repeated measures literature (Ferron, Dailey, & Yi, 2002; Gomez, 2005; Kesselman, Algina, Kowalchuk, &Wolfinger, 1999). Commonly used fit indices include deviance statistics, AIC, or BIC (Ferron, Dailey, & Yi, 2002; Singer & Willet, 2003). In general, the studies found that neither of the fit indices uniformly selected the correct level-one error structure.

The second method utilized in selecting the level-one covariance structure *a priori*. Sivo, Fan, and Witta (2005) wanted to investigate the degree to which autocorrelation in its various forms (AR, MA, ARMA) biases the estimates in GC modeling. They found that unmodeled autocorrelation could lead to biased results. The authors further suggested that more work was necessary with small samples to determine if the conclusions of the study would still hold. Additionally, Kwok, West, and Green (2007) conducted a similar study which sought to investigate various forms of misspecification (underspecification, overspecification, and general misspecification) of the level-one error structure. In the study, the authors concluded that it is, at times; best to adopt a slightly overspecified level-one error structure. The study investigated the appropriateness of these methods to small samples or single-case designs.

Based on the conditions used in these prior works, this study examined both methods: fit indices (post hoc selection) and *a priori* selection of the level-one error structure. More specifically, there were seven levels of this model specification factor: (1) *a priori* selection of independent, ID; (2) *a priori* selection of first-order autoregressive, AR(1); (3) *a priori* selection of first-order autoregressive first-order moving average, ARMA(1,1); (4) *post hoc* selection of either ID, AR(1) or ARMA(1,1) based on the AIC; (5) *post hoc* selection of either ID, AR(1) or ARMA(1,1) based on the AICC; (6) *post hoc* selection of either ID, AR(1) or ARMA(1,1) based on the BIC; and (7) *post hoc* selection of either ID, AR(1) or ARMA(1,1) based on the likelihood ratio test. The AIC selected model was the model that produced the lowest AIC value. The AICC selected model was the model that produced the lowest AICC value. Lastly, the BIC selected model was the model that produced the lowest BIC value. The lowest value was used as the model selected by each fit index, no fixed criteria, such as difference of at least 0.5 or 1, was used. To determine the model selected by the LRT, differences in fit at the .05 level were examined among the three models. If no significant differences were found the ID specification was selected. If the AR(1) produced significantly better fit than the ID, but not the ARMA(1,1), then AR(1) model was selected. Finally, if the ARMA(1,1) had significantly better fit than the other two previously discussed models, then the ARMA(1,1) was selected. Type I error was not controlled for the various tests, for example, for the ARMA model(1, 1) to be selected, two tests would have had to be rejected at the .05 level. Table 3 below shows a sample of the possible scenarios for the log likelihood ratio test (LRT). This led to a total of seven levels for the analysis factor.

Table 3
A Sample of the Tests that were used to Select the Correct Model for the LRT

ID vs. AR	ID vs. ARMA	AR vs. ARMA	Model Selected
Reject	Reject	Reject	ARMA
Reject	Reject	FTR	AR
Reject	FTR	FTR	AR
FTR	FTR	FTR	ID

The seven experimental factors that were previously described fall into three generic categories. The first three factors are design factors. Factors 4, 5, and 6 represent the data factors. Lastly, factor 7 is the analysis factor.

Sample

Crossing two levels for the error variances with the two levels of the fixed effects with five levels of covariance structures that were generated led to 20 conditions, which were then crossed with the 8 combinations of the number of studies included in the meta-analysis, the number of participants in each study, and the number of observations in the series length. For each of the design and data factors (160 conditions), 5000 simulated data sets were generated using SAS IML (SAS Institute, Inc., 2008). The use of 5000 data sets led to a standard error of .0003 for the confidence interval estimate at the .95 confidence level, which was an appropriate level of precision for this study.

Data Generation

The data were generated using Van den Noortgate's three-level model for the meta-analysis of single-case data. Equations 28-36 below represent the model formally,

Level 1 Equation:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}*phase_{ijk} + \beta_{2jk}*time_{ijk} + \beta_{3jk}*phase_{ijk}*time'_{ijk} + e_{ijk}, \quad e_{ijk} \sim N(0, \Sigma_e) \quad (28)$$

Level 2 Equations:

$$\beta_{0jk} = \theta_{00k} + u_{0jk} \quad (29)$$
$$\beta_{1jk} = \theta_{10k} + u_{1jk} \quad (30)$$
$$\beta_{2jk} = \theta_{20k} + u_{2jk} \quad (31)$$
$$\beta_{3jk} = \theta_{30k} + u_{3jk} \quad (32)$$
$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim N(0, \Sigma_u)$$

Level 3 Equations:

$$\theta_{00k} = \square_{000} + v_{00k} \quad (33)$$
$$\theta_{10k} = \square_{100} + v_{10k} \quad (34)$$
$$\theta_{20k} = \square_{200} + v_{20k} \quad (35)$$
$$\theta_{30k} = \square_{300} + v_{30k} \quad (36)$$
$$\begin{bmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v_{30k} \end{bmatrix} \sim N(0, \Sigma_v)$$

In Equation 28, the value of the dependent variable on measurement occasion *i* for subject *j* of study *k* (Y_{ijk}) is regressed on a dummy variable phase, that equals one if measurement occasion *i* occurs in the treatment phase, otherwise it is equal to 0 in the baseline phase. The score on the dependent variable on measurement occasion *i* for subject *j* of study *k* is Y_{ijk}. The *phase* variable is a dummy-coded indicator reflecting whether the observation is in the baseline phase ($phase_{ijk} = 0$) or the intervention phase ($phase_{ijk} = 1$). The coefficient for the *time* variable represents the slope during the baseline phase, $time_{ijk}$,and the coefficient for the interaction term, $phase_{ijk}*time'_{ijk}$, reflects the difference between the baseline and intervention phases' slopes. The *time* variable can be centered in a variety of ways which may be helpful for interpreting the model parameters (Baek et al., 2013; Van den Noortgate, & Ferron, 2013).

Specifically, for this study, the *time* variable is uncentered for the baseline phase's slope such that for the first measurement occasion in the baseline phase, $time_{ijk} = 0$. However, the *time′* variable in the interaction term, $phase_{ijk}*time'_{ijk}$is coded such that $time'_{ijk} = 0$ for the first observation in the intervention phase. Therefore, the expected score during the baseline phase if extended one observation into treatment is equal to β_{0jk} ; and the expected score of the treatment phase at this same point in time is β_{1jk} higher. As a result, β_{1jk} refers to the immediate level change associated with an intervention (Huitema & McKean, 2000). β_{2jk} represents the linear trend during the baseline phase, and the linear trend for the treatment phase was β_{3jk} higher. In other words, β_{3jk} indicates the effect of the intervention on the trend that is the difference between the baseline phase and treatment phase slopes. Both β_{1jk} and β_{3jk} are needed to fully describe the treatment effect. At the second level, Equations 29-32 describe the new regression equations for the variation over subjects. Equation 29 describes that the baseline performance for subject j from study k equals an overall baseline for study k plus some random deviation. Similarly, Equations 30-32 indicate the variation of the treatment effect, linear trend in baseline, and the effect of the intervention on linear trend, respectively, over subjects from the same study. The next set of equations can be thought of similarly as the second level equations.

At the third level, the variation across subjects is modeled using Equations 33-36. Equation 33 represents the baseline mean for study *k* as the overall baseline across all of the studies plus some random deviation. The same is modeled by Equations 34-36 for the variation of the treatment effect, linear trend in baseline, and the effect of the intervention on linear trend, respectively, across studies. It should be noted that errors on levels 2 and 3 are typically assumed to be normally distributed and have a mean of 0 and a variance of 1.0, however for this study the variances were as previously discussed. There was no covariance in the errors between levels and between errors at level-two and level-three.

The within-person error is modeled by e_{ijk} (σ^2_e represent the variance of e_{ijk}). Errors from the within-person, or level-one error structure, were generated using the ARMASIM function in SAS Version 9.3 (SAS Institute, 2008). There were five different level-one error specifications. The first represented an independent or ID level-one error structure with normal distribution, *N*(0, 1). The second generated the first order autoregressive level-one error structure, AR(1), autocorrelation coefficient of .2 or .4. Lastly, the first-order autoregressive first-order moving average, ARMA(1,1), level-one error structure was modeled, with both an autocorrelation parameter and moving average coefficient of .2 and .4, aligned with the values that were used for the autocorrelation parameter. The three equations (37-39) below represent the three different level-one error specifications, ID, AR(1), and ARMA (1,1), respectively. The data simulation was checked by examining the matrices produced at each stage. A small number of data sets were simulated to ensure that data specifications are accurate. The data set was analyzed and then reviewed to ensure that parameter estimates are close to expected estimates.

$$e_{ijk} = e_t \quad (37)$$

$$e_{ijk} = .3\, y_{(t-1)} + e_t \quad (38)$$

$$e_{ijk} = .3\, y_{(t-1)} - .3e_{t-1} + e_t \quad (39)$$

Analysis of Each Simulated Data Set

Model Specification

Each simulated data set was analyzed to provide results for each of the seven levels of the analysis factor. Moreover, each simulated data set was analyzed once using an ID specification of the level-one error structure, once using an AR(1) specification, and once using an ARMA(1,1) specification. In each case the three-level model was estimated using restricted maximum likelihood (REML) via PROC MIXED with the Kenward-Roger degrees of freedom method in SAS version 9.3 (SAS Institute Inc., 2008). The confidence interval for the variance components were constructed using the Satterthwaite approximation, which is the default in SAS 9.3.

Table 4 below represents the combinations for the covariance structure that was used in data generation versus the model that was used to analyze the data set.

Table 4
The Combination for the Type of Level-one Error Structure Generated and the Model Specification

	Model Specification		
Data Generated	ID	AR(1)	ARMA(1, 1)
ID	Correct	Over	Over
AR(1)	Under	Correct	Over
ARMA(1,1)	Under	Under	Correct

For example, the first row represents a data set that was generated assuming an uncorrelated level-one error structure, $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$. The data set was then analyzed using three distinct model specifications: the correct specification and two overspecifications-- a first-order autoregressive, AR(1) and a first-order autoregressive, moving average model, ARMA (1,1). For further illustration, the next row represents that the data were generated using the first-order autoregressive model. That data were then analyzed using the same three distinct models. However, underspecification was the uncorrelated error structure, correctly specified was represented by AR(1), and the first-order autoregressive moving average model demonstrated overspecification.

The results were further examined to determine which fit index correctly specified the level-one error structure. More specifically, variables were created to keep track of the accuracy of the fit indices in terms of selection of the level-one error structure. Additionally, the AIC, AICC, BIC, and LRT were examined, and this information was used to identify the model that would have been selected based on each of the fit indices. Moreover, the results from the three-model specifications that were previously mentioned were examined, in addition to the results that were produced by each of the various fit indices. This led to the seven levels of the analysis factor.

Table 5 below illustrates an example of three data sets and the hypothetical parameter estimates (true estimate, $\square_{100} = 2.0$) that would result from the three-model specifications and the estimates that would result from the models being selected by the fit indices. These indices may select the correct specification or one of the incorrect specifications. Bolded results represent the hypothetical true level-one error structure. According to the table below, the results for the first data set illustrate the estimated fixed effects for each of the three level-one error structures (the first 3 columns of the table). The first row also includes results for each of the models selected by the various fit indices. Specifically, the AIC and AICC selected the model with $\mathbf{\Sigma}$ = ID; the BIC selected the model with $\mathbf{\Sigma}$ = ARMA(1,1); and finally, the LRT selected the correct model, $\mathbf{\Sigma}$ = AR(1). Similarly, for the second data set, the true error structure is ID. For that data set, all of the fit indices (AIC, AICC, BIC, and LRT) correctly selected the level-one error structure. Lastly, for the third data set, the true level-one error structure is ARMA(1,1), and only the LRT correctly selected the ARMA(1,1) model.

Table 5
An Example of Results from Three Simulated Data Sets for the Shift in Level Effect, $\square_{100}$.

	A priori Model Selection			Fit Indices Model Selection			
Data set	ID	AR(1)	ARMA(1,1)	AIC	AICC	BIC	LRT
1	1.95	**1.98**	1.92	1.95	1.95	1.92	**1.98**
2	**2.02**	1.97	1.95	**2.02**	**2.02**	**2.02**	**2.02**
3	1.93	1.99	**1.99**	1.98	1.93	1.93	**1.99**

Based on the model, the shift in level was modeled as a change in level between the baseline and the treatment phases with the fixed effect for the shift in level ($\square_{100}$) set to 2.0. The effect of the intervention on the trend was modeled as the change in slopes between the baseline and the treatment phases with the fixed effect for the intervention effect of the slopes ($\square_{110}$) set to 0.2. Estimates were also obtained for the moving average parameter, autocorrelation parameter, variance within participants, variance in baseline levels across participants and studies, variance in treatment effects (shift in level) across participants and studies, variance in baseline slopes across participants and studies, and variance in treatment effect on the trends (change in slope) across participants and studies.

Summary of analyses

First, the study investigated the accuracy of the fit indices in selecting the appropriate covariance structure when using the three-level model to meta-analyze single-case data. An indicator variable for each fit index was used to indicate whether or not the fit index correctly identified the correct level-one error structure. This variable was averaged across all 5000 data sets to obtain the proportion of times that each fit index correctly identified the error structure. Next, the accuracy of the fixed effects and the variance components were analyzed using the following dependent variables: bias (see Equation 40 below) can be described as the difference between the known parameter and the estimated values from the model output, for two of the four fixed effects: **(**$\square_{000}$, $\square_{100}$, $\square_{010}$, $\square_{110}$**)** and all of the variance components for level-one (σ^2, ρ, and the moving average coefficient); level-two $\sigma^2_{u_1}, \sigma^2_{u_3}$); and level-three ($\sigma^2_{v_1}, \sigma^2_{v_3}$); this difference or deviation was then averaged across all 5000 data sets to obtain the average bias. Additionally, relative bias was calculated for the parameters whose known values were other than 1 and did have levels of the factor that included 0. The RMSE represented the square root of the sum of the squared residuals; this was calculated by squaring the deviations between the estimated parameter and the true parameter, taking the average across all 5000 data sets, and finally the square root to obtain the RMSE (see Equation 41).

Confidence interval coverage (the proportion of the confidence intervals at the .95 level that contained the true parameter estimates for both the fixed effects and the variance components) was again tracked with an indicator variable that determined whether or not the parameter estimate fell within the confidence interval range. This indicator was then averaged across the 5000 data sets to obtain the proportion of confidence intervals that contained the true parameter estimates. Similarly, confidence interval width (the average difference between the upper and the lower limits of the 95% confidence intervals for both the fixed effects and the variance components) was calculated for each of the 5000 data sets. The width was averaged across the 5000 data sets to obtain the average confidence interval width. These outcomes were computed for each of the 1120 conditions obtained by crossing the 160 data and design conditions with the seven levels of the analysis factor. The percentage of non-convergence was also computed for each of the 160 conditions.

$$\text{bias} = \frac{\sum_{n=1}^{5000}\left(\hat{\gamma}_{1j} - \gamma_{1j}\right)}{5000} \tag{40}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{5000}\left(\hat{\gamma}_{1j} - \gamma_{1j}\right)^2}{5000}} \tag{41}$$

Analyses to Examine Relationships between Design Factors and Outcome Variables

Research Question One

The evaluation of the accuracy of the fit indices to correctly select the level-one error structure was addressed by examining the percentage of time that each fit index appropriately guided the researcher to the correct error structure. For example, data were generated using a first-order autoregressive level-one error structure. The model was then run using all three of the aforementioned level-one error specifications, i.e. two misspecified models (uncorrelated and first-order moving average) and the correct model (a first-order autoregressive). The results of the three models was compared to estimate the proportion of times the AIC correctly identified the model (smallest AIC value), the proportion of times the BIC identified the model (smallest BIC value) the proportion of times the AICC identified the model (the smallest AICC value), and the proportion of times the likelihood ratio test (LRT) correctly identified the model (LRT statistically significant at $\alpha = .05$). Also, the results of the model selected by the fit indices was used for comparison of model selected by *fit indices*, or post hoc selection versus model selected *a priori*. The model selected a priori is the model the models that are run without the use of the fit indices. For example, the ID, AR(1), and ARMA (1,1) model was run to see how these results compared to the fit-index selected models. Additionally, to examine which design factors explained the variability that was found among the fit indices, general linear modeling (GLM) was used. The GLM model allowed the examination of the variability of the fit index selection as a function of the independent variables. The models were built with the criteria of discerning the effects whose eta-squared values were at least .06 or greater. The effect size, eta-squared (η^2) was calculated to determine the proportion of variability associated with, or explained by each effect. Using Cohen's (1988) criteria for eta-squared, a small effect was described as $\eta^2 = .01$; a medium effect size $\eta^2 = .06$; and a large effect as having $\eta^2 = .14$ or greater. Each model initially included only the main effects. Whether or not more complex parameters were added to the model was based upon the amount of variability that the first model explained. Specifically, if the model that contained only main effects explained a significant proportion of the variability, then neither two-way nor three-way interactions were added. However, if the fixed-effects only model did not explain the minimum 94% of the variability, then two-way interactions were included in the model. Finally, if the model still did not explain 94% of the variability, then more complex interactions were added to the model, such as three-way and four-way interactions until at least the 94% of the variability had been explained. If a medium effect was found ($\eta^2 >= .06$), further follow-up analyses were conducted. A comparison of the means were done using line graphs to expound on the relationship between the different levels of that factor (e.g. number of studies included in the meta-analysis) and the variability of the outcome (e.g. AIC selected models).

Research Question Two – Research Question Five

The remainder of the research questions was addressed similarly. The evaluation of the outcome of interests and the parameter estimates from the three level model used to meta-analyze single case data were addressed by examining box-and-whisker plots to illustrate the distribution of the parameter estimates. Furthermore, to examine which design factors explained the variability that was found among the parameter estimates, general linear modeling (GLM) was used. The models were built with the criteria of discerning the effects whose eta-squared values were at least .06 or greater. The effect size, eta-squared (η^2) were calculated to determine the proportion of variability associated with, or explained by each effect. Using Cohen's (1988) criteria for eta-squared, a small effect was described as $\eta^2 = .01$; a medium effect size $\eta^2 = .06$; and a large effect as having $\eta^2 = .14$ or greater. Each model only included main effects. Whether or not more complex parameters were added to the model was based on the amount of variability that the first model explained. Specifically, if the model that contained only main effects explained a significant proportion of the variability, then neither two-way nor three-way interactions were added. However, if the fixed-effects only model did not explain the minimum variability, then two-way interactions were included in the model. Finally, if the model still did not explain the minimum variability, then more complex interactions were added to the model, such as three-way and four-way interactions until a substantial amount of variability had been explained. If a medium effect was found ($\eta^2 = .06$), further follow-up analyses were conducted. A comparison of the means were done using line graphs to expound on the relationship between the different levels of that factor (e.g. number of studies included in the meta-analysis) and the variability of the outcome (e.g. confidence interval coverage).

Chapter Summary

This chapter outlined the methods for this study as well as described the purpose, research questions, design, and simulation conditions. The data generation methods, analytical procedures, and outcome measures have also been discussed. The goal of this chapter was not only to illustrate and to build upon previous work that had been done with the use of meta-analyzing single case research using the three-level model, but also to extend this work by investigating various level-one error structure misspecifications and the use of fit indices to select the level one error structure. This Monte Carlo work will not only guide methodologists, but can also guide single-case researchers when determining intervention effectiveness in the presence of correlated level-one error structures.

CHAPTER FOUR: RESULTS

The chapter displays the results for the six research questions in sequential order. The chapter begins with a detailed description of how the results were obtained. First, the accuracy of the fit indices were examined. This is followed by the section that presents the outcomes of interest (bias and RMSE, confidence interval coverage and width, and Type I error and power) as related to the fixed effects, while the second half presents similar outcome of interests (bias, RMSE, confidence interval coverage and width) as related to the variance components. The following research questions were addressed:

1. To what extent do fit indices (log likelihood ratio test, AIC, adjusted AIC, and BIC) correctly identify level-one covariance structure when using a three-level meta-analytic single-case model?
2. To what extent are the **fixed effects** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures), and analysis factors (form of specification)?
3. To what extent are **confidence interval width and coverage for the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study,

series length per primary study), data factors (variances of the error terms, covariance structures), and analysis factors (form of specification)?

4. To what extent are the **Type I error and power for the test of the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures), and analysis factors (form of specification)?
5. To what extent are the **variance component** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, and covariance structures), and analysis factors (form of specification)?
6. To what extent are **confidence interval width and coverage for the variance components** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures), and analysis factors (form of specification)?

There were 1120 conditions simulated in this Monte Carlo study using the seven design factors. These factors included the 1) the number of primary studies per meta-analysis (10 and 30); 2) the number of participants per primary study (4 and 8); 3) the series length or number of observations per participant (10 and 20); 4) variances of the error terms (most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at higher levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]); 5) the level for the fixed effects (

0 or [2 for the shift in level and .2 for the shift in slope]); 6) the level of autocorrelation and the moving average parameter, respectively: [(0,0), (.2, 0), (.4,0), (.2, .2), (.4, .4)]; and 7) the form of model specification [i.e. ID, AR(1), ARMA (1,1)], and error structure selected by AIC, AICC, BIC, and the LRT. Finally, this resulted in a 2x2x2x2x2x5x7 factorial design.

First, the extent to which each fit index could correctly identify the covariance structure was evaluated. This question involved analyzing the proportion of times that each fit index correctly selected the appropriate level-one error structure. This was accomplished by first, looking at the box plots which illustrated the distribution of the correctly specified models across the four fit indices. Then, GLM models were run to explain the variability in the various proportions for the fit indices, with the dependent variable representing the correct proportion and the independent variables were the design factors in the study (the number of observations or series length, the number of participants, the number of studies to be included in the meta-analysis, the variances of the error terms, the level of the treatment effect, the level of the autocorrelation and the moving average parameters, and the type of fit index).

Next, the dependent variables or outcomes of interest (bias, RMSE, confidence interval coverage and width, type I error, and power) were analyzed for the fixed effects and the outcomes of interest (bias, RMSE, confidence interval coverage and width) for the variance components were evaluated. In order to compare the outcomes of different sizes, the relative bias was calculated for all of the outcomes where the parameter value was not equal to 1.0. The results of the study was then analyzed using PROC GLM in SAS to assess the relationship between the independent variables or outcomes of interest (bias, RMSE, confidence interval coverage and width, type I error, and power) and the dependent variables or the design factors for the simulation study (the number of primary studies per meta-analysis, the number of participants per primary study, the series length or number of observations per participant, variances of the error terms, the level for the fixed effects, the level of autocorrelation and the moving average parameter, and the form of model specification). In other words the outcomes of interest were the dependent variables and the design factors were modeled as the independent variables.

These models were built with the intention of finding medium effects or larger (whose eta-squared values were equal to or greater than .06). The effect size, the eta-squared values (η^2) was calculated to measure the degree of association between the dependent variable and the main effects, or interactions, if necessary of the independent variables or the study's design factors. Eta-squared is the proportion of variability in the outcome measure that is explained or associated with each of the effects in the simulation study. The formula is included below and can be described as the ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}):

$$\eta^2 = SS_{effect} / SS_{total}$$

The calculated η^2 values were compared to Cohen's (1988) standards for interpreting eta-squared values with a small effect as $\eta^2 = .01$; a medium effect as $\eta^2 = .06$; and a large effect as $\eta^2 = .14$ or greater. Each model was first created using a main-effects only model. If this model did not explain at least 94% of the total variability, then higher order interactions (second-order interactions were added, then third-order interactions, and so on) were included in the model until at least 94% of the variability was explained. However, if the model explained at least 94%, then it was known that no interaction effects were necessary. If at least a medium effect was found, then line graphs or a box plots were created to further investigate the association between the outcomes of interest and the study's design factors. All of the samples converged.

Overall Correct Fit Index Identification

The first question involved the accuracy of the fit indices, specifically the proportion of times that each fit index correctly identified the correct level-one error structure.

ID Model

The box plot below (see Figure 4) illustrates the proportion of times that each fit index correctly identified the correct error structure, when the structure was the ID model. The largest mean value (*M* = 0.85, *SD* = 0.02), indicating that the ID model was correctly identified most often by the log likelihood ratio test (LRT). The smallest mean value (*M* = 0.67, *SD* = 0.03) for the proportion of times that the ID model was correctly selected was for the model selected by the AIC fit index.

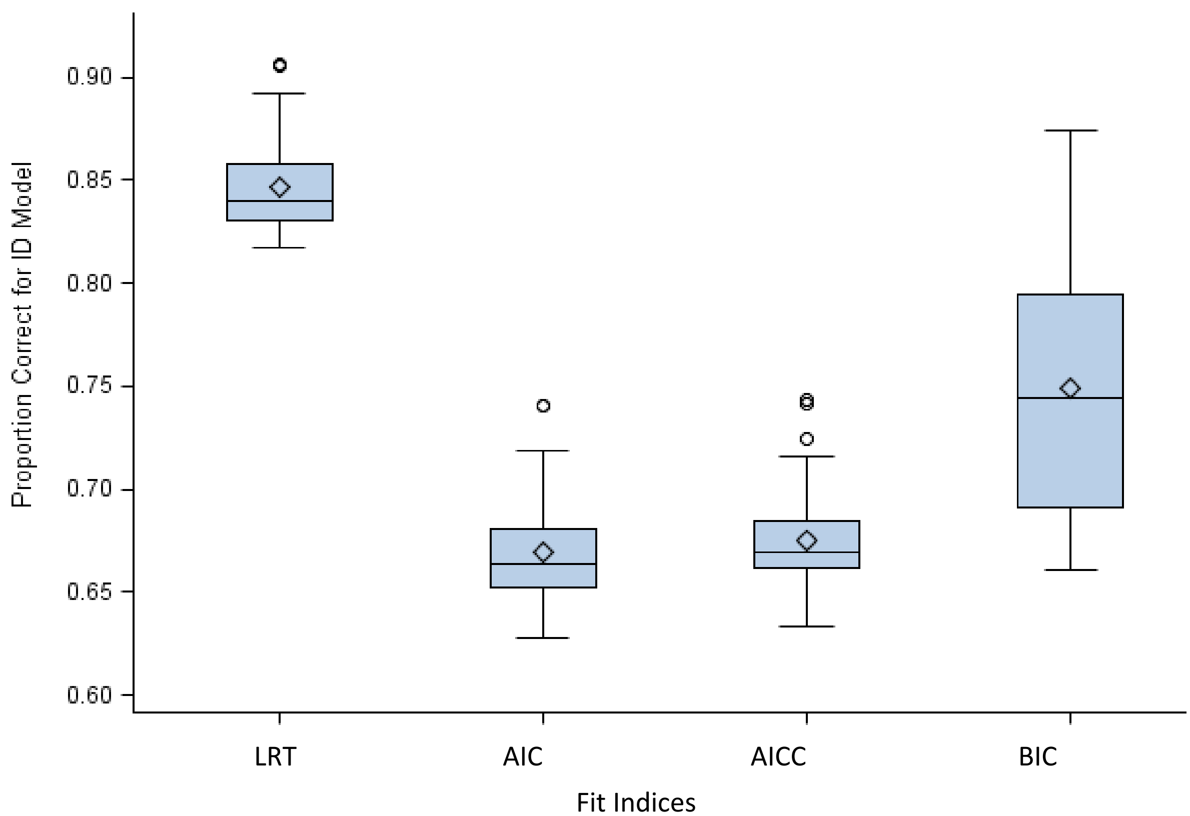

Figure 4. Boxplot representing the distribution of proportion correct for each fit index when the model specified is the ID model.

To further explore the variability in the proportion of times that each fit index correctly identified the ID model, a GLM model was run across these seven design factors (series length, number of participants in the study, number of primary studies included in the meta-analysis, variances of the error terms, the level of the fixed effects, and the type of fit index). The model, including two-way interactions, which explained 99% of the total variability revealed that the interaction effect between the number of primary studies included in the meta-analysis and the type of fit index (η^2 = .064) used for model selection, met the aforementioned criteria for having at least a medium effect.

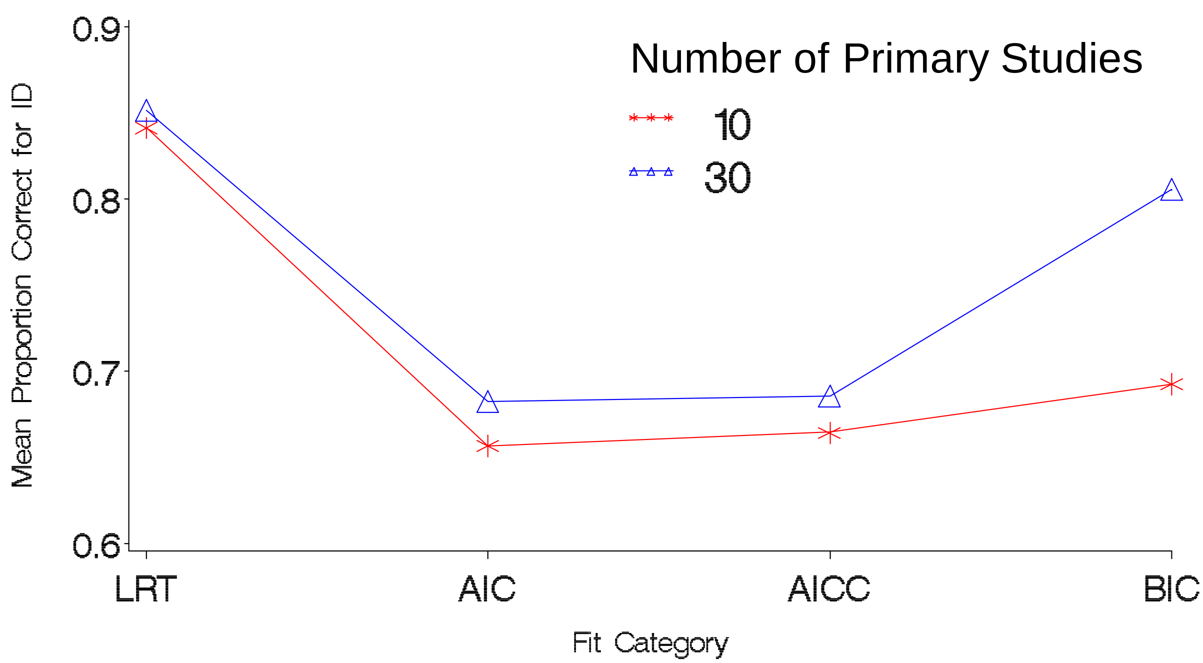

Figure 5. Line graphs illustrating the mean proportion of times that the ID model was correctly specified for each of the four fit indices.

Line graphs were then created to further examine the relationship between the proportion of times that the ID model was correctly selected with the interaction effect including the number of primary studies included in the meta-analysis and the fit index used for selection. Figure 5 shows that the when the number of primary studies is increased (from 10 to 30), then the proportion of the times that the ID model is correctly specified is increased across all fit indices. However, this increase is not identical across all fit indices. More specifically, when the number of studies is increased from 10 (*M* = 0.69, *SD* = 0.02) to 30 (*M* = 0.81, *SD* = 0.04), the mean increase in the proportion of times that the ID model is correctly selected is greatly improved by the BIC fit index. The least mean increase for the proportion of times that the fit index correctly selects the ID model is seen for the LRT. For the models selected by the LRT, when the number of primary studies is 30 (*M* = 0.85, *SD* = 0.03), the mean proportion is slightly greater than when the number of primary studies included in the meta-analysis is 10 (*M*= 0.84, *SD* = 0.02).

First-order Autoregressive Model

The box plot in Figure 6 below depicts the proportion of time that the various fit indices correctly selected the first-order autoregressive, AR(1), model. The AIC index had the least mean proportion of correctly identifying the AR(1) model (*M* = 0.80, *SD* = 0.06). The greatest proportion for the correct selection of the AR(1) model was for the log likelihood ratio test, LRT, *M* = 0.90, *SD* = 0.09. To further explore the variability observed in the box plots, GLM models were created.

The model, including three-way interactions explained 98% of the total variability, and affirmed three effects that were significant: the series length or number of observations (η^2 = .10), the number of primary studies included in meta-analysis (η^2 = .16), and the type of fit index used for selection (η^2 = .24). Furthermore, additional plots were then used to further examine the relationship of the mean proportion of correctly identifying the AR(1) model with these other effects.

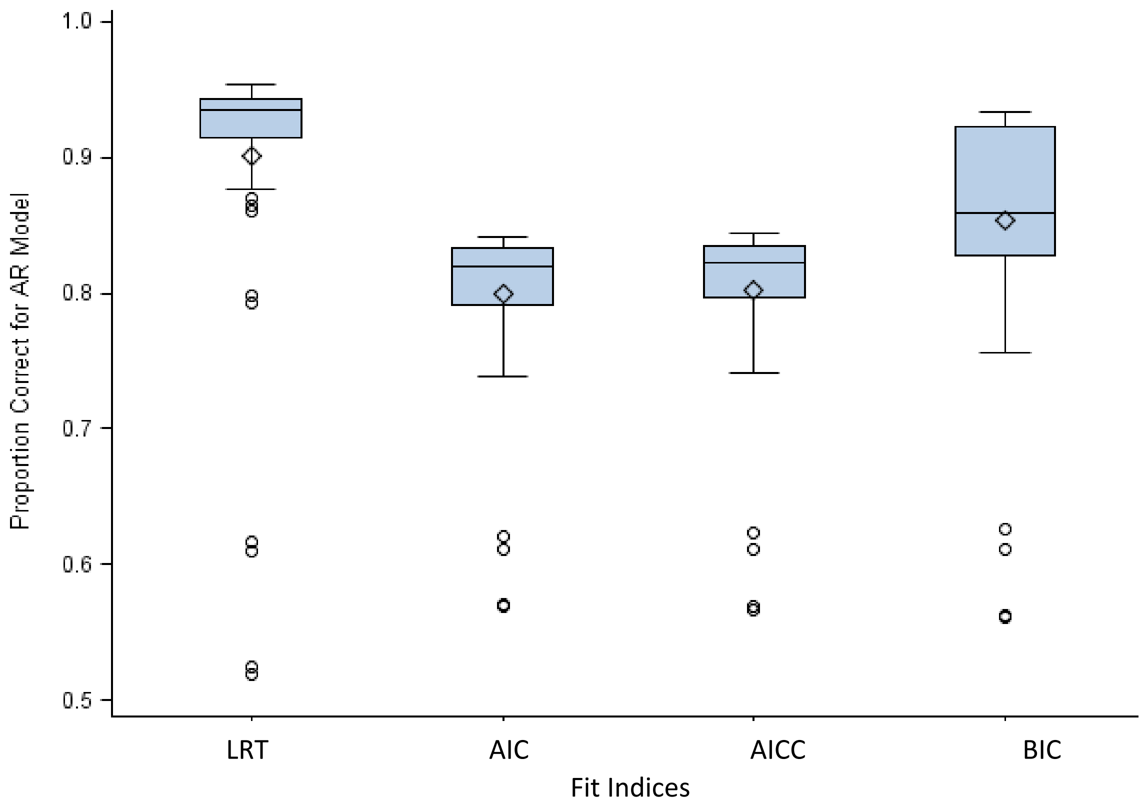

Figure 6. Boxplot representing the distribution of proportion correct for each fit index when the model specified is the AR model.

Figure 7 below shows the direct relationship between the proportion of time for selecting the AR model correctly across the levels for series length. Moreover, the graph depicts that the as the series length increases from 10 to 20 then the proportion for correctly identifying the AR model also increases from a mean of 0.81 (*SD* = 0.10) to a mean of 0.87 (*SD* = 0.05). The variability is also decreased with increased number of observations or a greater series length.

Next, the association between the mean proportion of correctly selecting the AR(1) model and the number of primary studies is depicted in Figure 8 below. The means for the proportion of correctly selecting the AR(1) model is shown above (see Figure 8, pg. 90) across the levels for the number of primary studies to be included in the meta-analysis. The graph illustrates that as the number of studies increases from 10 (*M* = 0.81, *SD* = 0.10) to 30 (M = 0.87 , SD = 0.05), the proportion of times for correct identification of the AR(1) model also increases.

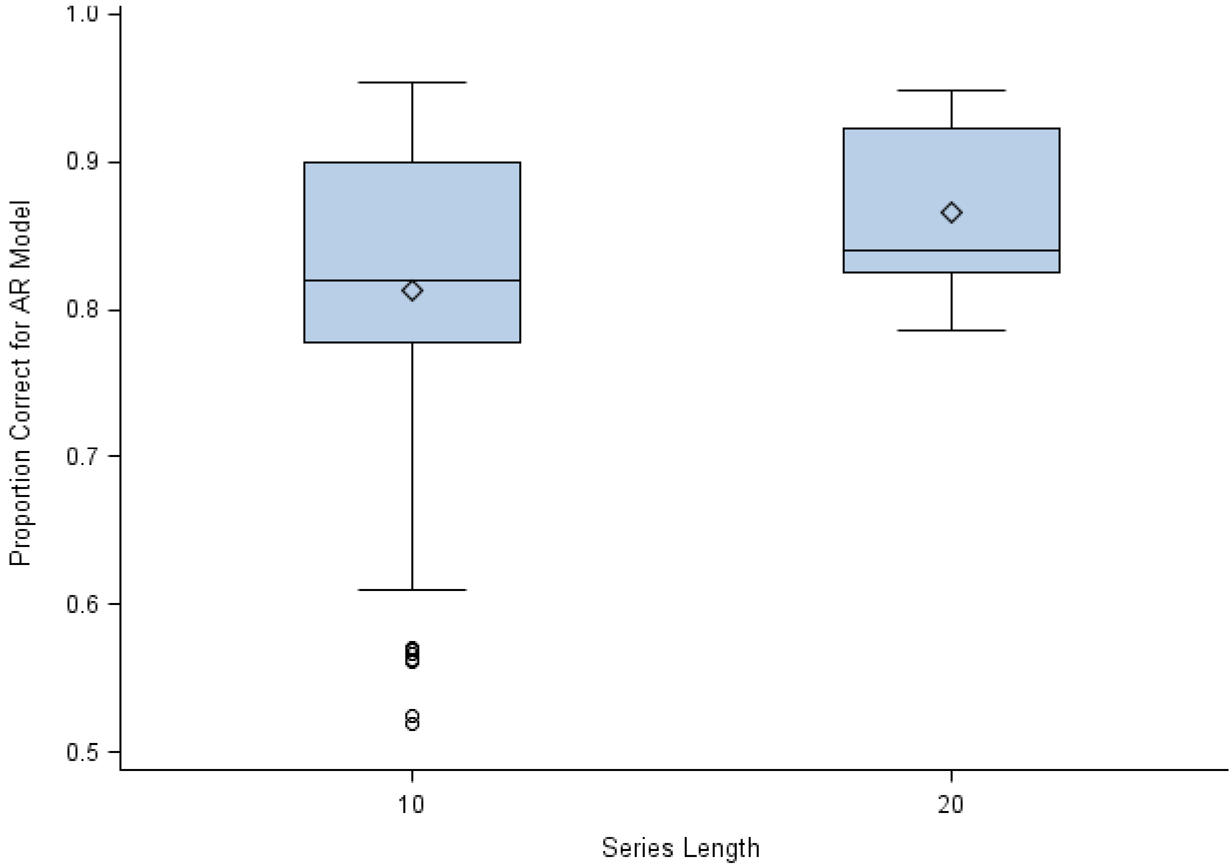

Figure 7. Box plots illustrating the distribution for the proportion of the AR model that was correctly specified with series length.

Next, the association between the mean proportion of correctly selecting the AR(1) model and the number of primary studies is depicted in Figure 8 below. The means for the proportion of correctly selecting the AR(1) model is shown above (see Figure 8) across the levels for the number of primary studies to be included in the meta-analysis. The graph illustrates that as the number of studies increases from 10 (*M* = 0.81, *SD* = 0.10) to 30 (M = 0.87 , SD = 0.05), the proportion of times for correct identification of the AR(1) model also increases.

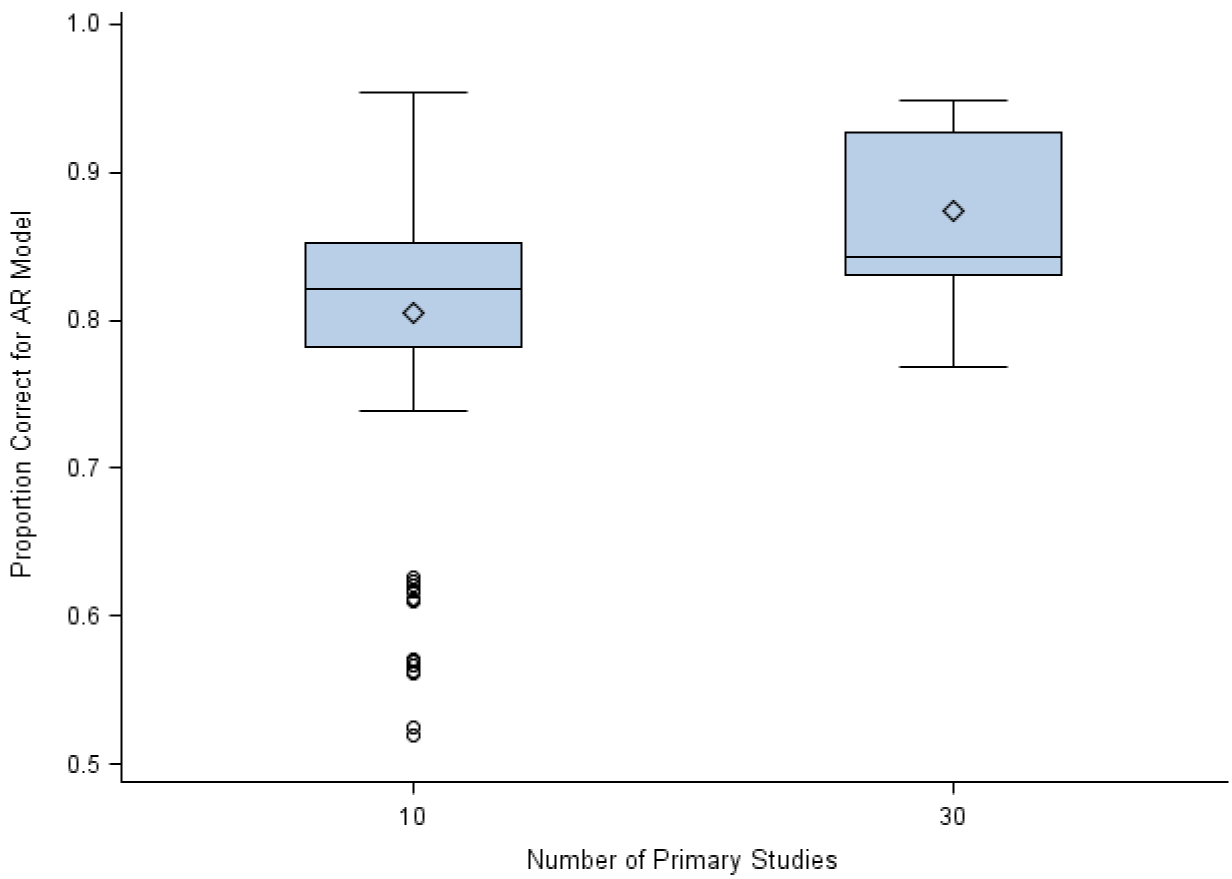

Figure 8. Box plots illustrating association of the mean proportion of correctly selecting the AR model and the number of primary studies included in the meta-analysis.

First-order Autoregressive Moving Average model

The distribution for the proportion of times that the first-order autoregressive moving average model, ARMA (1, 1) was correctly identified is shown in Figure 9 below. The box plot below illustrates the distribution of the mean proportion of times that the first-order autoregressive moving average model was correctly identified by the various fit indices.

Although there appears to be some variability among the fit indices, none of the fit indices correctly identified the ARMA model more than 20% of the times. The greatest mean proportion for correctly selecting the ARMA(1,1) model was found when the fit index used was the AIC (*M* = 0.19, *SD* = 0.02). The fit index with the smallest mean proportion times for correctly identifying the ARMA(1,1) model (*M* = 0.07, *SD* = 0.12) was the LRT. To further explore the variability, GLM models were run.

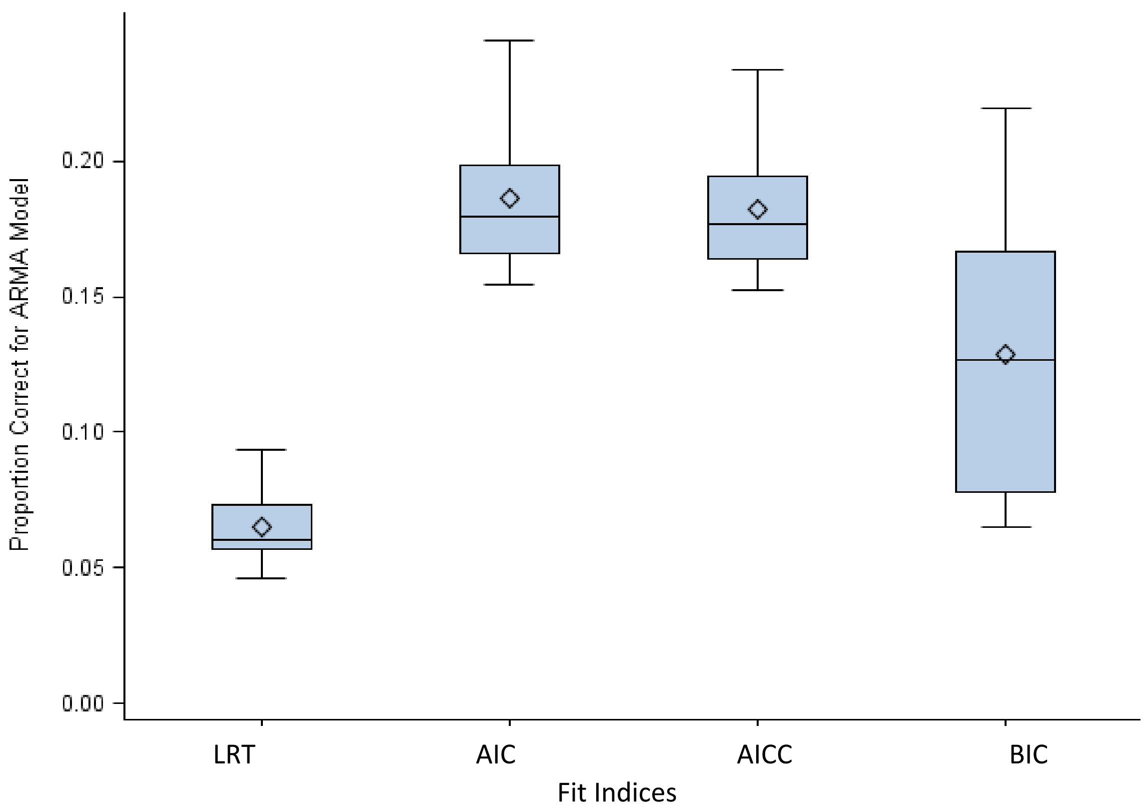

Figure 9. Box plot illustrating the distribution of the mean proportion of times that the first-order autoregressive moving average model was correctly identified by the various fit indices.

The model, including two-way interactions, explained 98% of the total variability, and indicated that the interaction effect between the number of primary studies to be included in the meta-analysis and the type of fit index (η^2 = .08) constituted a medium effect. The relationship for the mean proportion of correctly identifying the first-order autoregressive moving average model and the interaction effect between the number of primary studies to be included in the meta-analysis and the type of fit index is depicted by the line graph below (see Figure 10).

The graph shows that the effect of the number of primary studies to be included in the meta-analysis depends on the type of fit index used to select the correct model. Specifically, as the number of primary studies is increased from 10 (*M* = 0.17, *SD* = 0.03) to 30 (*M* = 0.08, *SD* = 0.02), the greatest decrease is observed for the model selected by BIC. As the number of primary studies to be included in the meta-analysis is increased from 10 (M= 0. 07, SD = 0.01) to 30 (M= 0. 06, SD = 0.01), the smallest impact is seen for the LRT.

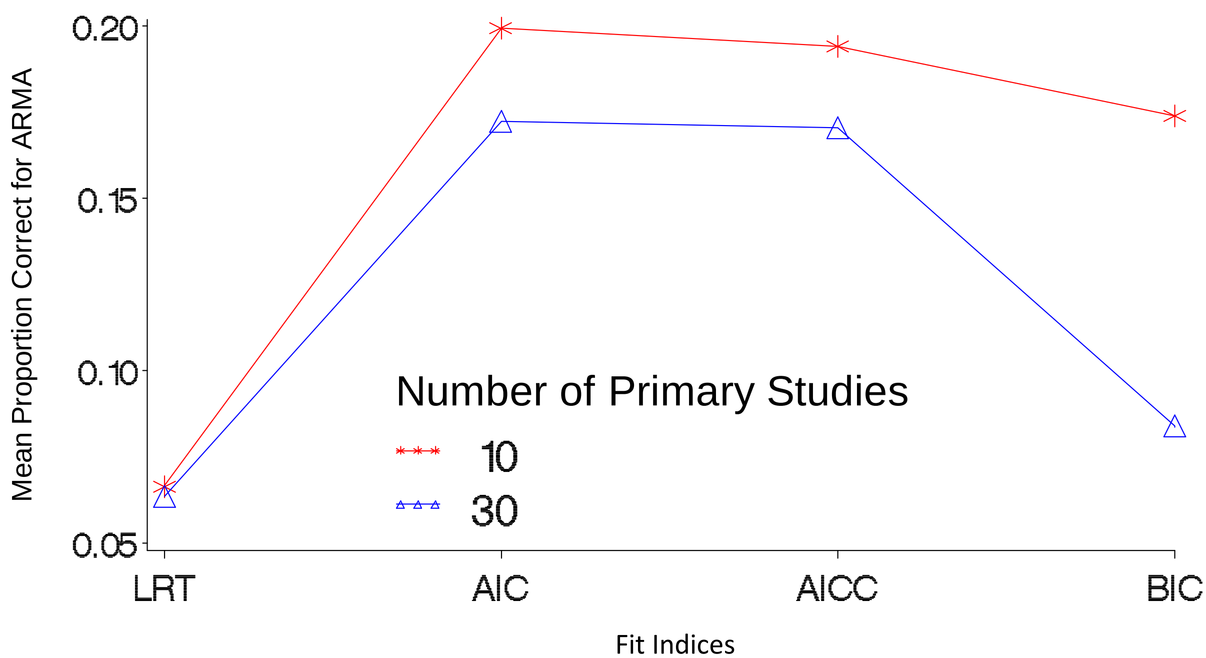

Figure 10. Line graph illustrating the association between the mean proportion for correctly selecting the ARMA (1,1) model and the interaction effect of between the number of primary studies to be included in the meta-analysis and the type of fit index.

Fixed Effects

The second question referred to the bias and the RMSE associated with the fixed effects as a function of the seven factors used in this Monte Carlo study. The third question described the extent to which confidence interval coverage and width of the fixed effects varied as a function of the seven design factors. Furthermore, question four involved the degree to which there was variability in the power and Type I error of the fixed effects as a function of the study's design factors. The percentage of non-convergence was not an issue in the study, convergence rates were all 1.

Bias

The distribution of bias values for the fixed effect for the shift in level (phase) and the interaction effect (shift in slopes) is shown in Figure 11 and Figure 12 below, respectively.

Overall average treatment effect for the phase (shift in level). The average bias values were close to 0 across all seven models with little to no variation, the eta-squared value for the type of model was .000099588. Specifically, the average bias was the smallest in magnitude (*M* = -.00001, *SD* = .0043) for the model selected by the log likelihood ratio test (LRT) and largest was for the first-order autoregressive moving average model, ARMA (1,1), (*M* = -.00009, *SD* = .0044). As indicated by the results above, there was little to no variability across the models for the average bias values. GLM models, including 5-way interactions, were run to see if any of the design factors had a significant effect, but none were found. Although the model explained 95% of the variability, none of the effects met the aforementioned criteria for a medium effect. Due to this finding, no further exploration was appropriate, and the variability that was observed in the bias values can be attributed to sampling error.

Overall average treatment effect for the interaction effect (shift in slopes). Similar results were found for the bias values in interaction effect (shift in slopes). The distribution of the bias values is shown in Figure 12. The average bias values were close to 0 across all seven models with again, little to no variation.

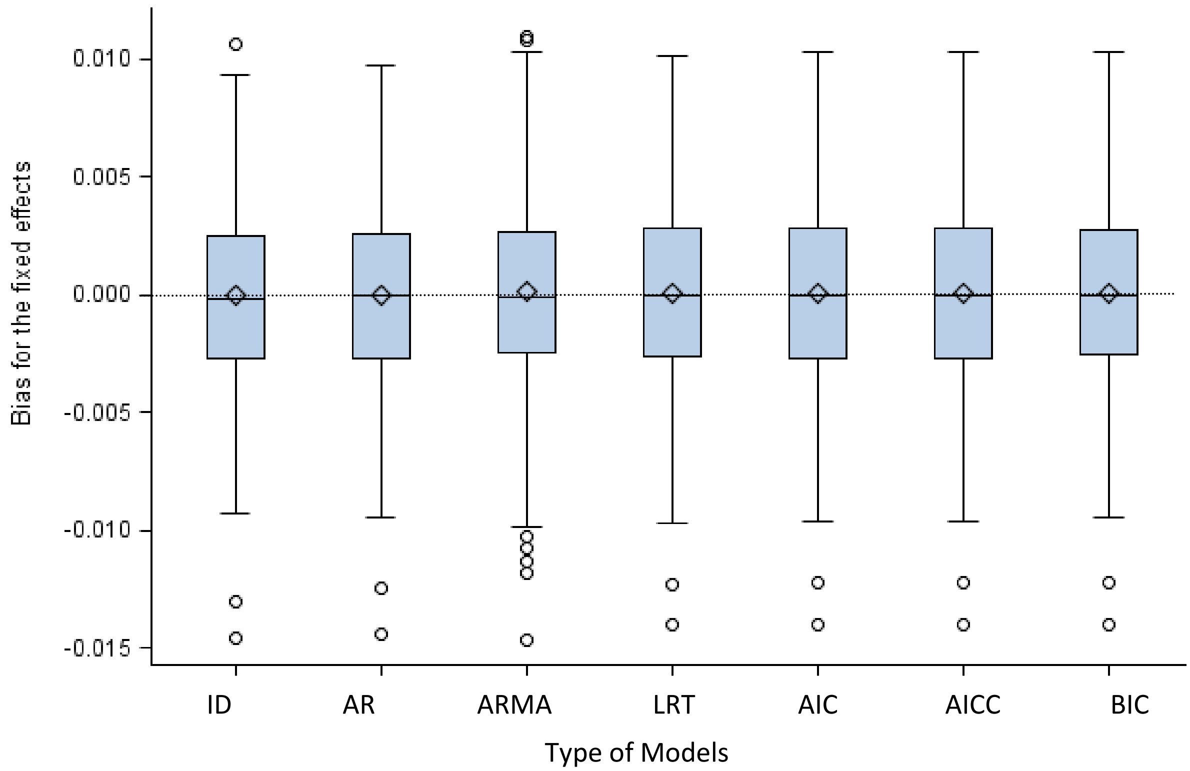

Figure 11. Box plots illustrating the distribution for the bias values for the phase effect (shift in level) across the 7 different models.

The type of model (η^2 = .000017531) explained very little of the variability in the bias values, again indicating similarity of the bias values across models. Specifically, the average bias was the smallest (*M* = -.000001250, *SD* = .0017) for the LRT model. The first-order autoregressive, AR(1), model had the largest average bias value (*M* = .000001250, *SD* = 0.0016). These values reveal very little to no bias present in the fixed effects. According to Hoogland and Boomsma (1998), parameter estimates are acceptable with relative bias values less than five percent. The bias values for the overall interaction effect are well below this criterion, therefore no further analyses was warranted. The bias illustrated in the figure below can be attributed to sampling error, and have minimal impact on the parameter estimates.

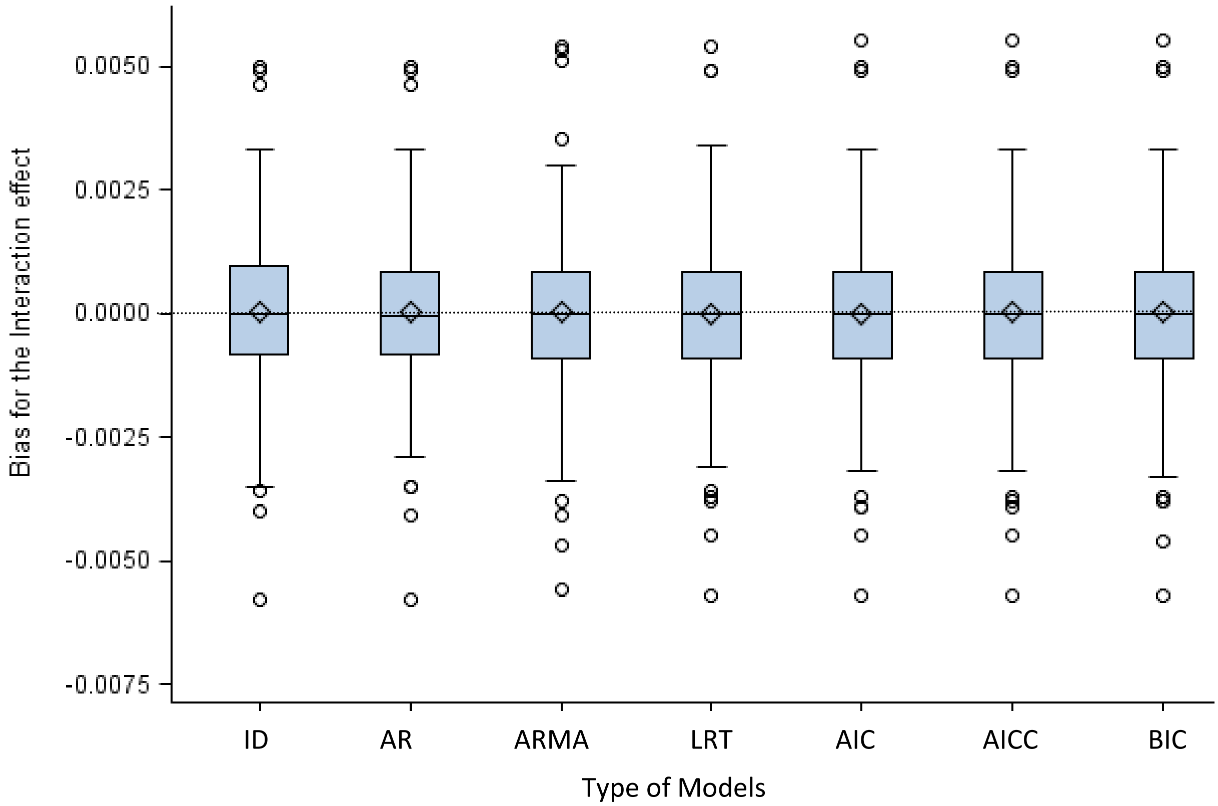

Figure 12. Box plots illustrating the distribution for the bias values for the interaction effect (shift in slopes) across the seven different models.

Root Mean Square Error (RMSE)

The distribution of the RMSE values for each of the intervention effects (shift in level and shift in slopes) is depicted across the seven models in Figures 13 and 16, respectively.

Overall average treatment effect for phase (shift in level). The average RMSE value for the treatment effect for phase (shift in level) was similar across the seven models; with the type of model explaining very little of the variability ($\eta^2 = .00004$). This small eta-squared value reinforced the noticeably small amount of variability across the seven models. The smallest

RMSE mean value was for the first-order autoregressive model (*M* = 0.32, *SD* = 0.13). The largest mean RMSE value (*M* = 0.32, *SD* = .13) for the phase effect was for the ID model.

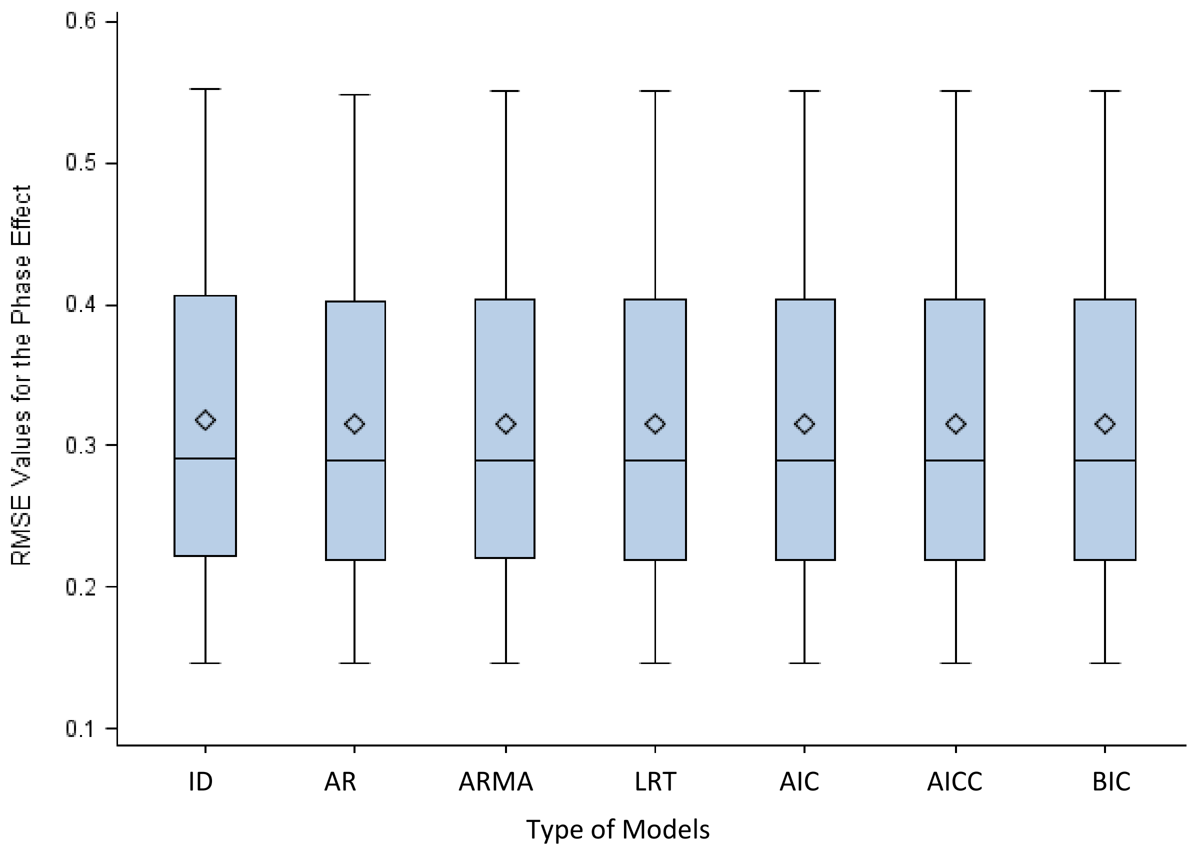

Figure 13. Box plots illustrating the distribution for the RMSE values for the phase (shift in level) across the seven models.

To further explore the variability in the RMSE values for the phase effect, a GLM model was created. The main effects only model explained over 96% of the variability and indicated that two of the design factors had at least a medium effect, number of primary studies included in the meta-analysis (η^2 = .45) and the variances of the error terms (η^2 = .49). The box plots below was used to represent the RMSE values as a function of the number of primary studies included in the meta-analysis. As illustrated in the graph (see Figure 14), the RMSE values decreased from a mean of 0.40 (*SD* = .11) to a mean of 0.23 (*SD* = 0.23) as the number of primary studies included in the meta-analysis increased from 10 to 30. There was also a noted difference in the variability of the RMSE values for the shift in level as the number of primary studies increased from 10 to 30.

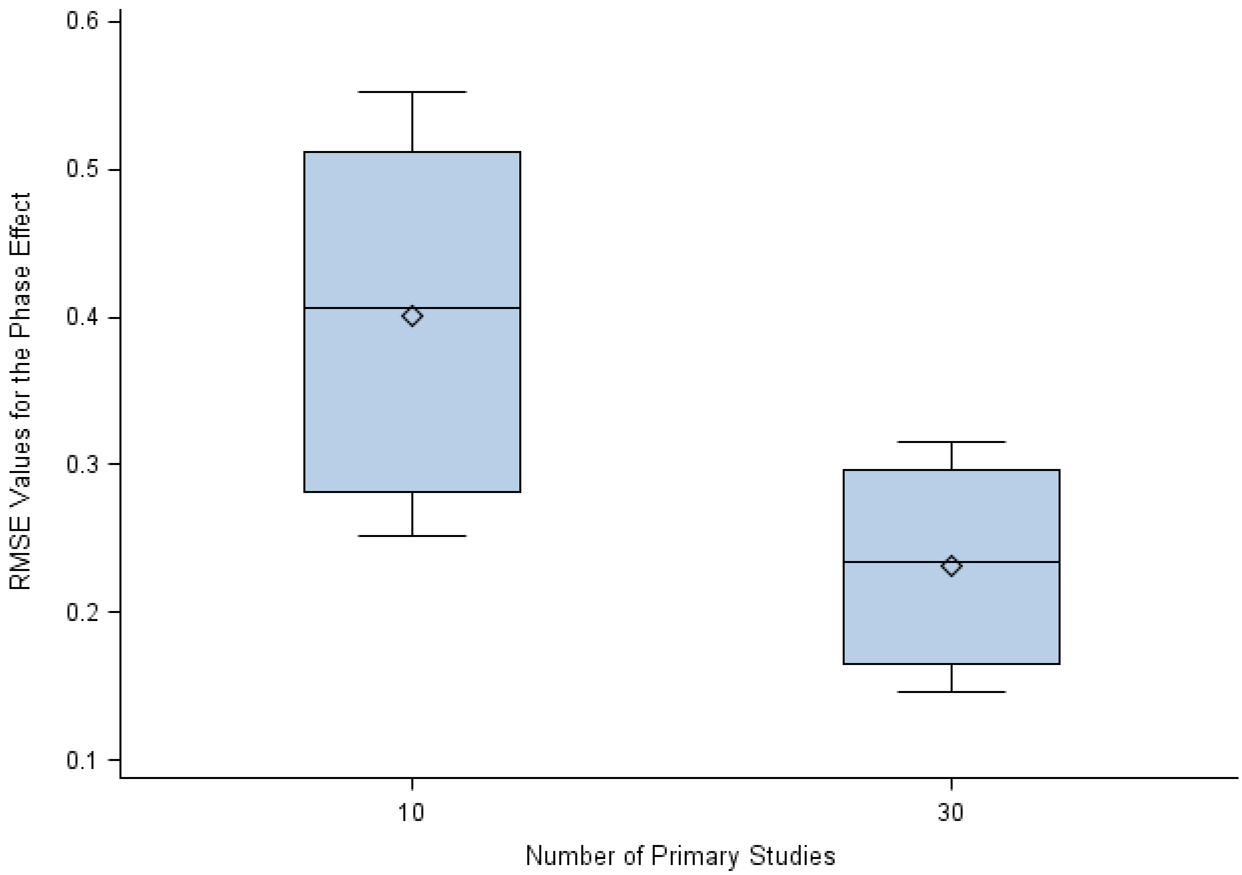

Figure 14. Box plot depicting the estimated RMSE values for the shift in slopes as a function of number of primary studies included in meta-analysis.

Similarly, a box plot was created to further analyze the relationship between the RMSE values and the variances of the error terms. Figure 15 below represents the relationship, moreover the figure portrays that as the variance shifts from most of the variance for the error terms at level-one to most of the variance at the upper levels, the RMSE mean values increase from 0.23 (*SD* = 0.06) to 0.40 (*SD* = 0.11). Note that the variance of the error terms for the residual variance was held constant at 1.0 across conditions. Thus, the conditions that have most of the variance shifted to the upper levels have more total variance. The variability in the RMSE values for the shift in level also tended to decrease with more variance in the upper levels as opposed to more variance at level-one.

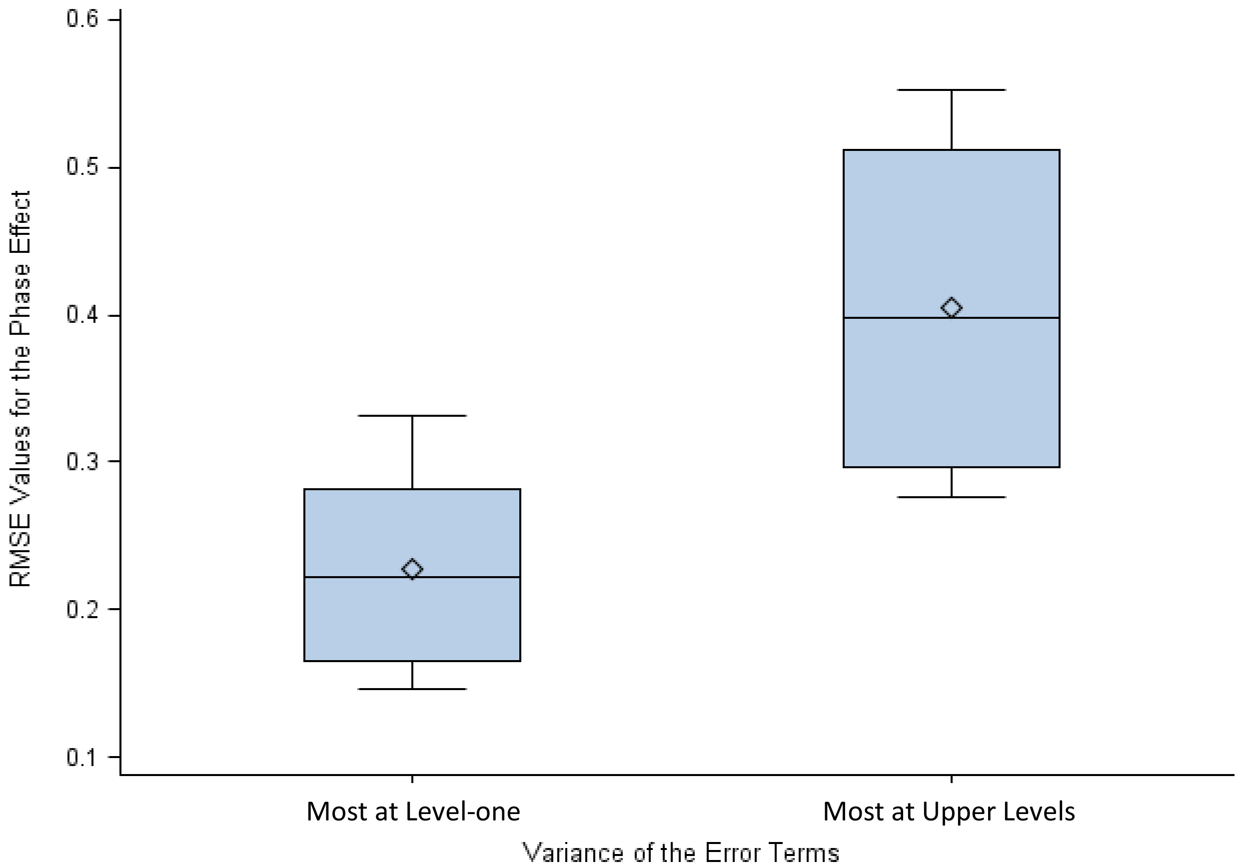

Figure 15. Box plot depicting the estimated RMSE values for the shift in level as a function of the variances of the error terms.

Overall average treatment effect for slopes (shift in slopes). The distribution of the RMSE values for the interaction effect is shown in Figure 16 below. The distribution was very similar across the seven models (η^2 = .00000), indicating little to no variability across models for the mean RMSE values. The smallest mean value (*M* = 0.10, *SD* = .042) for the first order autoregressive model (AR). The largest mean RMSE value can be observed for the ID model (*M* = 0.10, *SD* = .042).

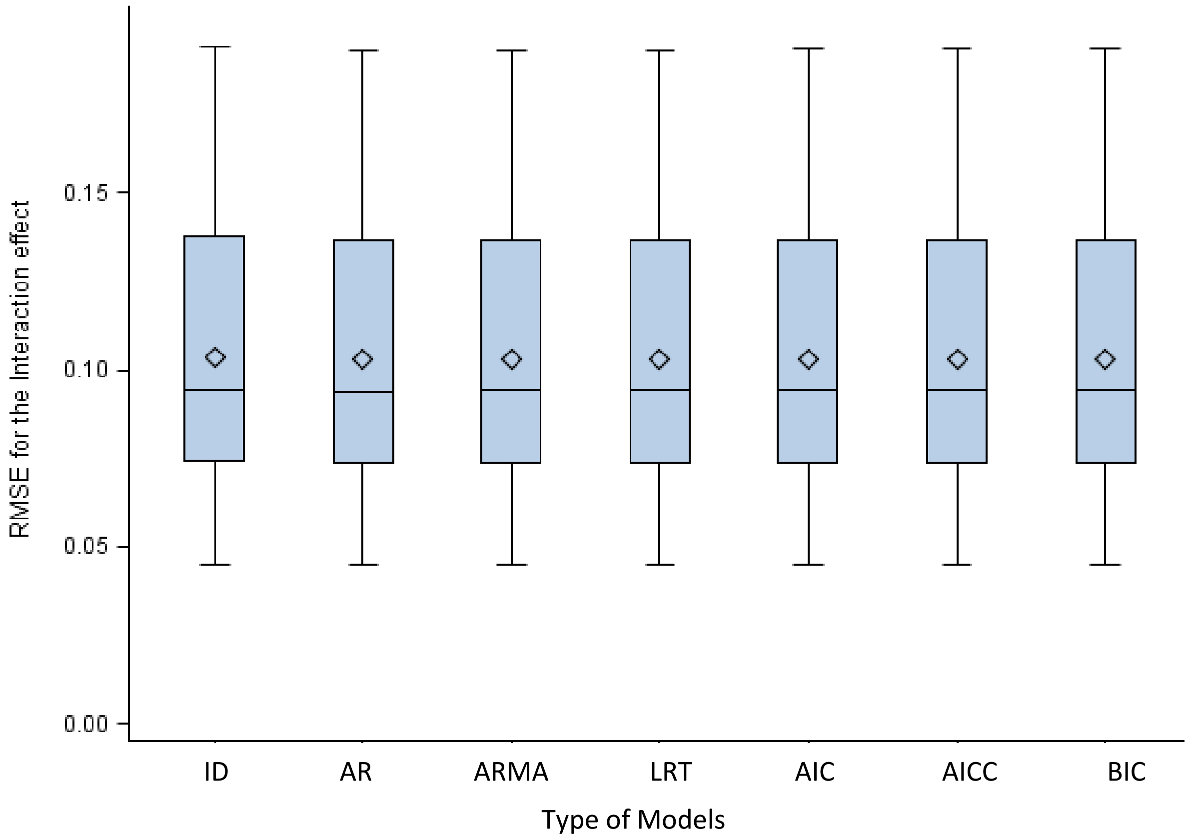

Figure 16. Box plots illustrating the distribution for the RMSE values for the interaction effect (shift in slopes) across the seven models.

Variation in the RMSE values were explored by modeling RMSE values across the seven design factors. The main effects only model explained 95.6% of the variability and revealed that only two of the design factors had at least a medium effect according to the aforementioned criteria. The means for the interaction effect for the RMSE values as a function of the number of primary studies (η^2 = .45) included in the meta-analysis are shown in Figure 17 below. As depicted in the figure, as the number of primary studies increased from 10 to 30, then the RMSE mean values decreased from 0.13 (*SD* = .04) to 0.08 (*SD* = 0.02).

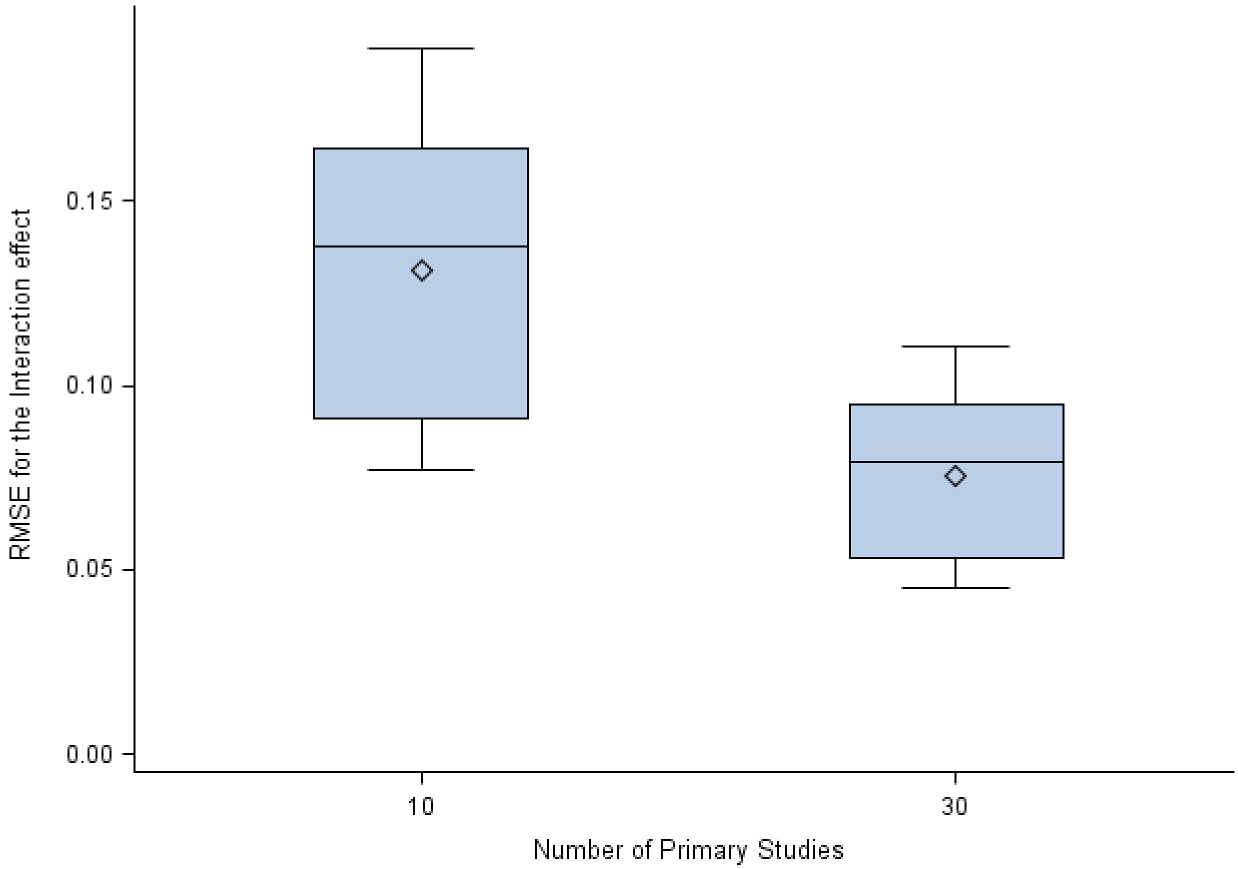

Figure 17. Box plot depicting the estimated RMSE values for the shift in slopes as a function of number of primary studies included in Meta-Analysis.

The second factor that had at least a medium effect was the variance for the error terms (η^2 = .45) in the main effects only model for the RMSE values for the interaction effect (shift in slopes). Figure 18 illustrates the relationship between the RMSE values and the variance for the error terms. Specifically, as the variance in the error terms shifted from most of the variance being at the level-one to most of the variance being at the upper levels, the RMSE mean values increased from 0.08 (*SD* = 0.02) to 0.13 (*SD* = 0.04).

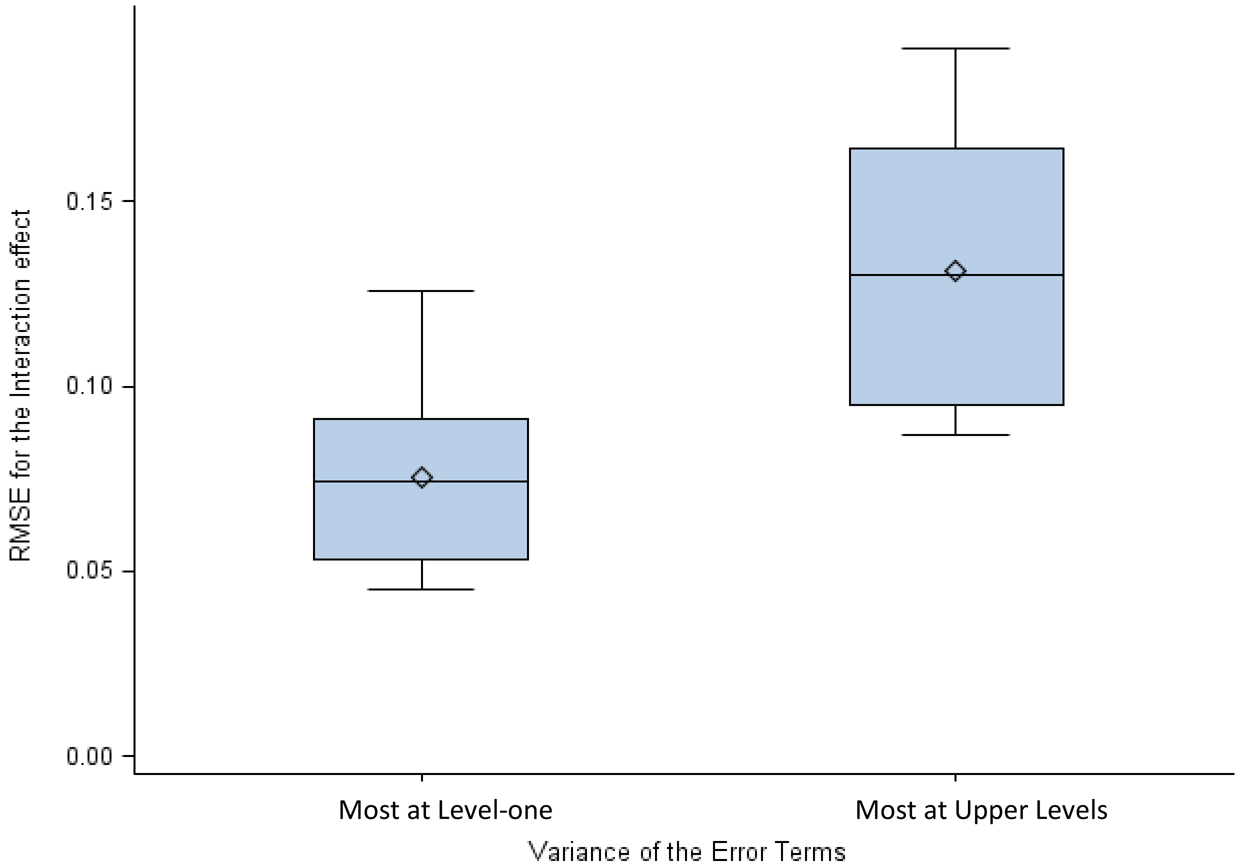

Figure 18. Box plot depicting the RMSE values for the shift in slopes as a function of the error variances.

Confidence Interval Coverage

The distribution of confidence interval coverage rates for each of the fixed effects is illustrated across the seven models in Figures 19 and 20 below.

Overall average treatment effect (shift in level). The mean confidence interval coverage rate was comparable across the seven models (see Figure 19 below), with means that are very close to the nominal value of 0.95. The type of model (η^2 = .008) explained little to none of the variability, which supported the small variation that was observed across models in the box plots. The smallest mean confidence interval coverage was for the ID model (*M* = 0.949, *SD* = 0.003); the largest mean confidence interval coverage (*M* = 0.951, *SD* = 0.003) was observed for the LRT model. The range for the interval coverage was 0.942 to 0.964, which falls within Bradley's (1978) criterion; therefore no further analyses were warranted.

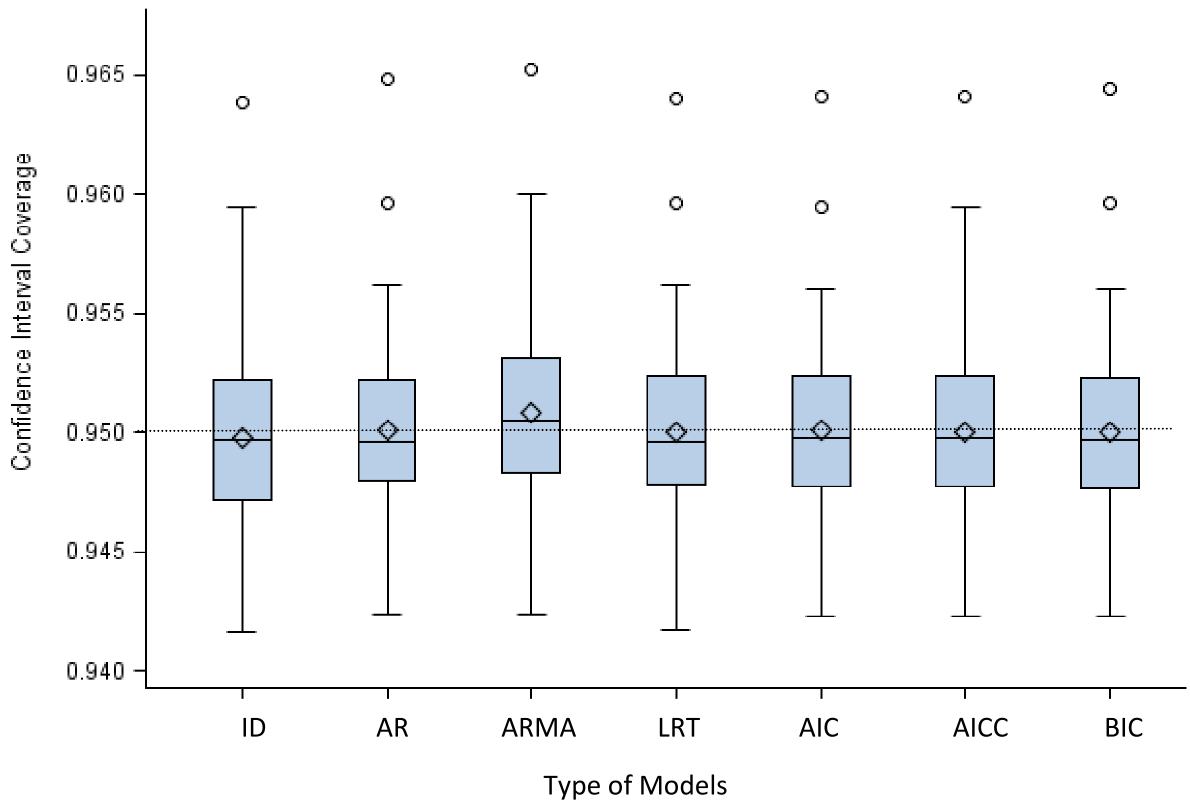

Figure 19. Box plots illustrating the distribution of confidence interval coverage rates for the phase effect (shift in level) across the seven different models.

Overall average treatment effect for slopes (shift in slopes). The average confidence interval coverage rates for the interaction effect, or the shift in slopes, were at the nominal level of .95, ranging from a mean of .948 (*SD* = .0033) for the ID model and .951 (*SD* = .0036) for the first-order autoregressive moving average model. The range for the interval coverage for the interaction effect was 0.939 to 0.964; this range falls within Bradley's (1978) criterion for acceptable coverage limits, therefore no further analyses was warranted in terms of explaining the variability by the study's design factors.

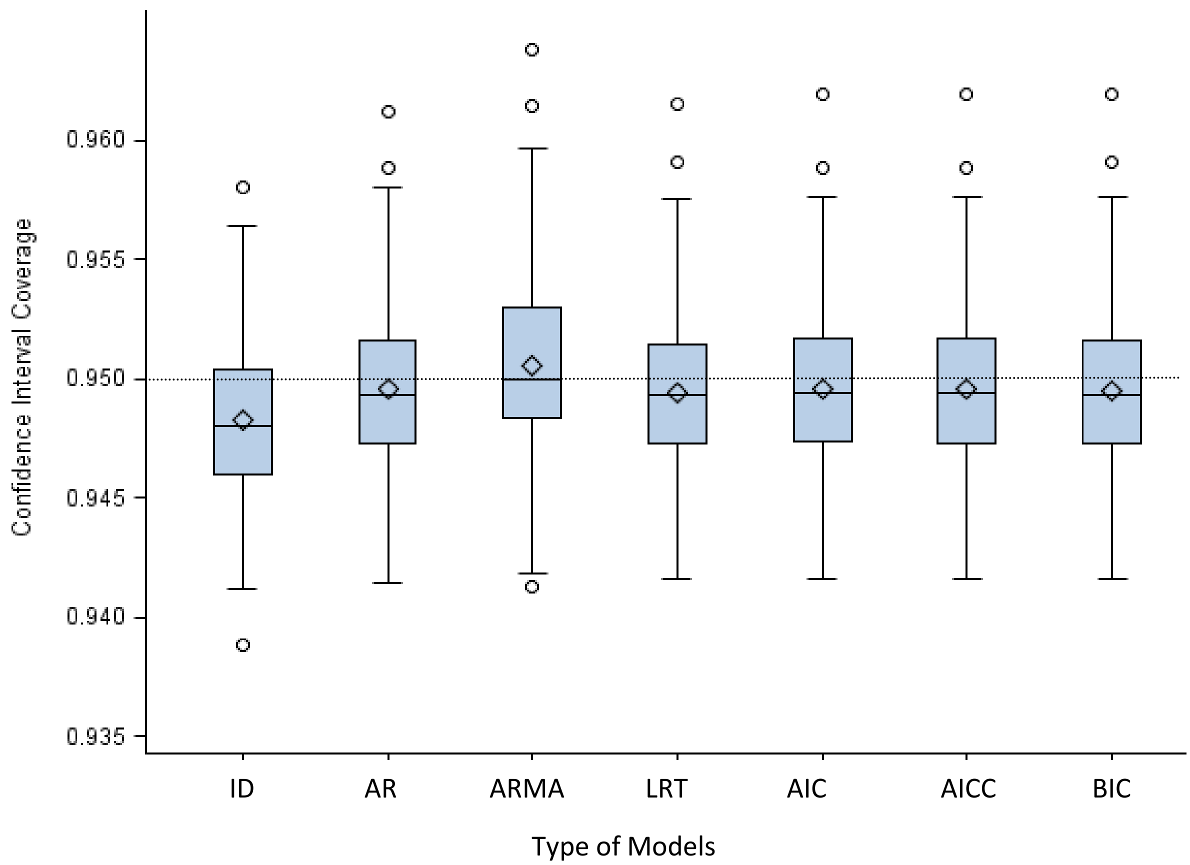

Figure 20. Boxplots illustrating the distribution for the confidence interval coverage for the interaction effect across the seven models.

Confidence Interval Width

The box plot depicting the distribution of the confidence interval width estimates for the two intervention effects (shift in level and shift in slopes) across the seven models are displayed below in Figures 21 and 25, respectively.

Overall average treatment effect for phase (shift in level). The average confidence interval width for the phase effect across the seven models was comparable. The mean confidence interval width was 1.35 (*SD* = .58) for first-order autoregressive model and the four fit-index selected models, however, for the ID model, which had a slightly larger mean width (*M* =1.36, *SD* = .58). The largest mean confidence interval width (*M* = 2.51, *SD* = 3.48) was for the first order autoregressive moving average model. Furthermore, the box plots reveal that there may be some variability within each model.

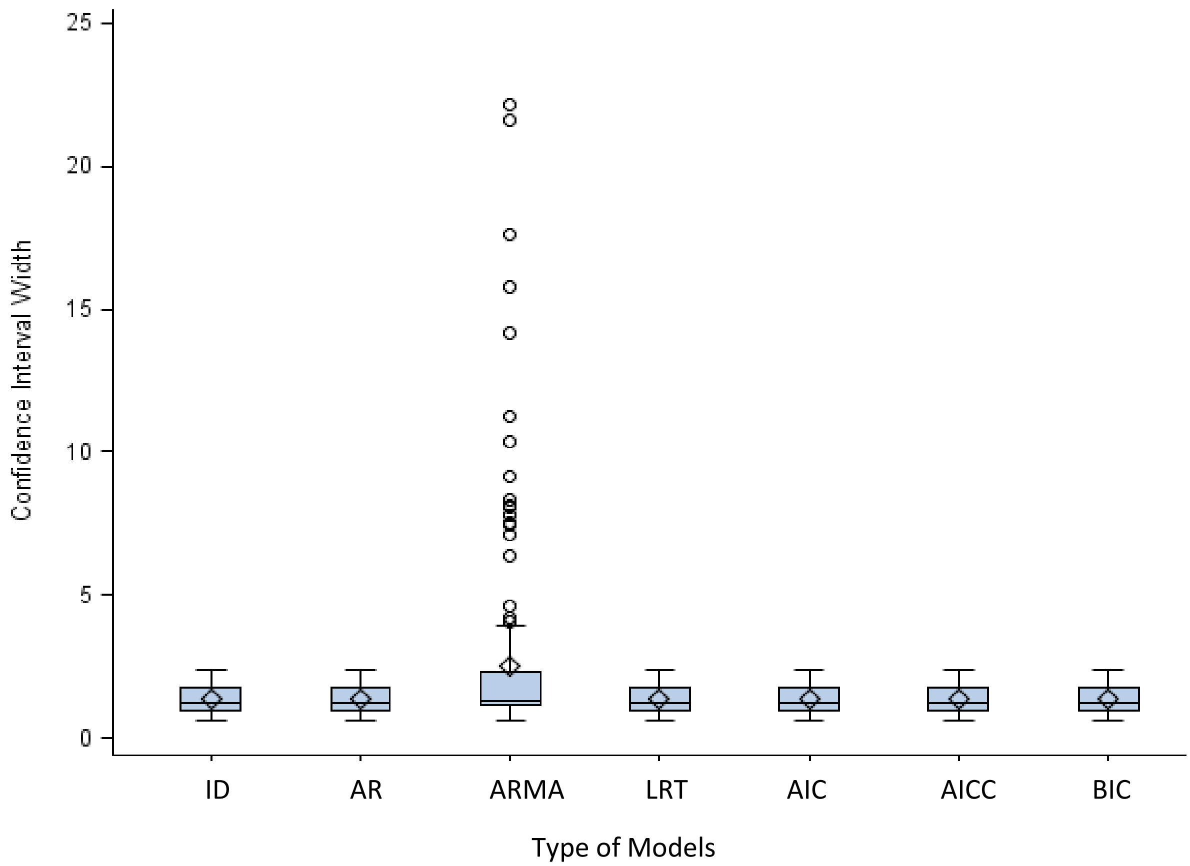

Figure 21. Boxplots to explore distribution of confidence interval width for the phase effect (shift in level) across all seven models.

GLM models were used to further investigate the variability by modeling confidence interval width as a function of the design factors. The model, including fourth-order interactions explained 96% of the variability, however only four effects met the aforementioned criteria as a medium effect: number of primary studies included in meta-analysis ($\eta^2 = .10$), the type of model ($\eta^2 = .08$), the variances of the error terms ($\eta^2 = .10$), and the interaction between autocorrelation parameter and type of model (η^2= .16). Graphs were created to further explore the relationship of confidence interval width with each of the significant effects.

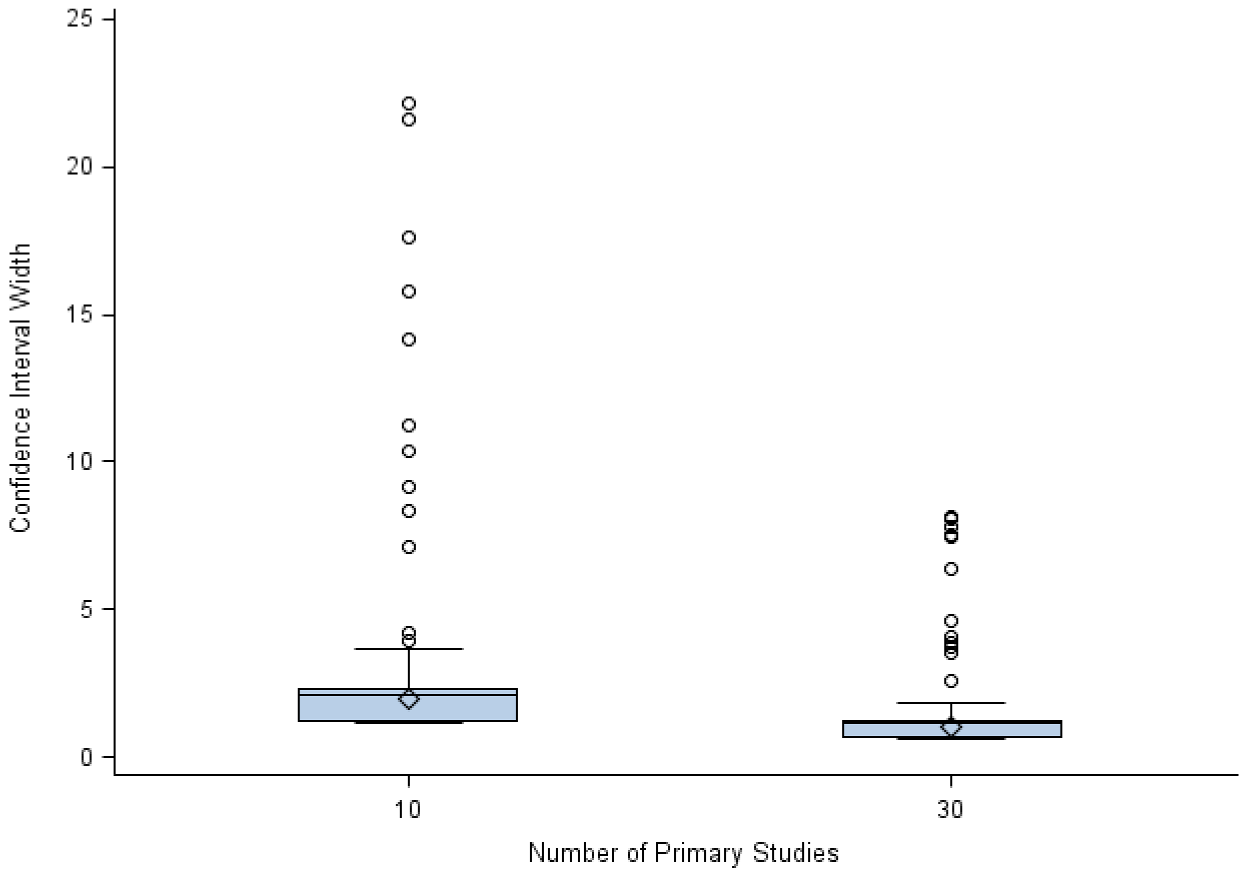

Figure 22. Box plot illustrating the confidence interval width for the phase effect (shift in level) as a function of the number of primary studies included in meta-analysis.

The first graph (see Figure 22 above) shows that as the number of primary studies included in the meta-analysis increased from 10 to 30, then the mean confidence interval width for the phase effect also decreased. Additionally, the plots again depict that there is a decrease in the variability of the interval coverage width for the phase effect (shift in level) when the number of primary studies is increased from 10 to 30. The variability also tended to decrease for the width of the phase effect when the number of primary studies increased.

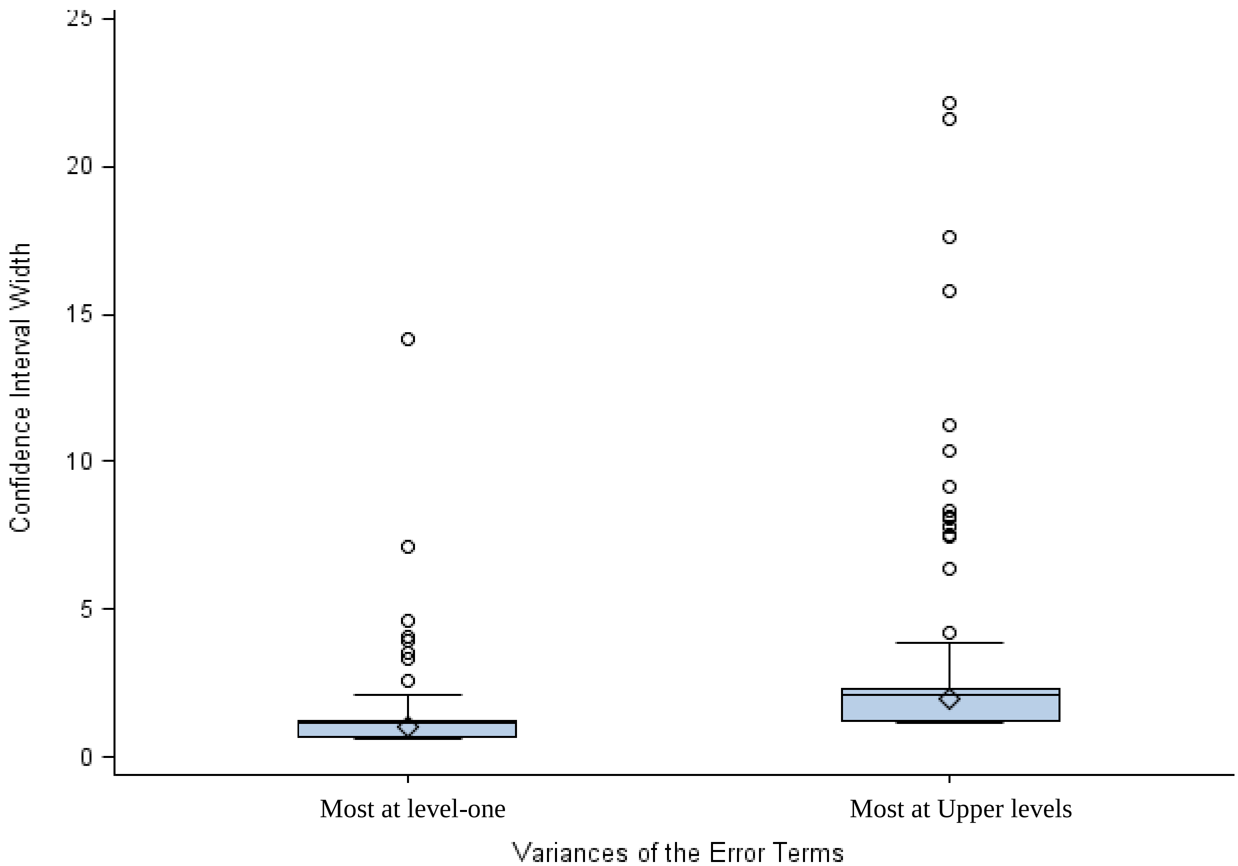

Figure 23. Graph illustrating the mean confidence interval width for the phase effect as a function of the variances of the error terms.

The above graph in Figure 23 depicts the relationship between the mean confidence interval width for the phase effect and the variances of the error terms. The box plot depicts that as the variances of the error terms shifted from most of the variance at level-one to most of the variance of the error terms at the upper levels, the mean interval width increased from 0.97 (*SD* = 0.31) to mean of 1.73 (*SD* = 0.53). However, there were also more outlying points when most of the variance shifted to the upper levels. Thus, when most of the variance shift from being at level-one to most of the variance being at the upper levels (more total variance), then the mean and variance tended to increase.

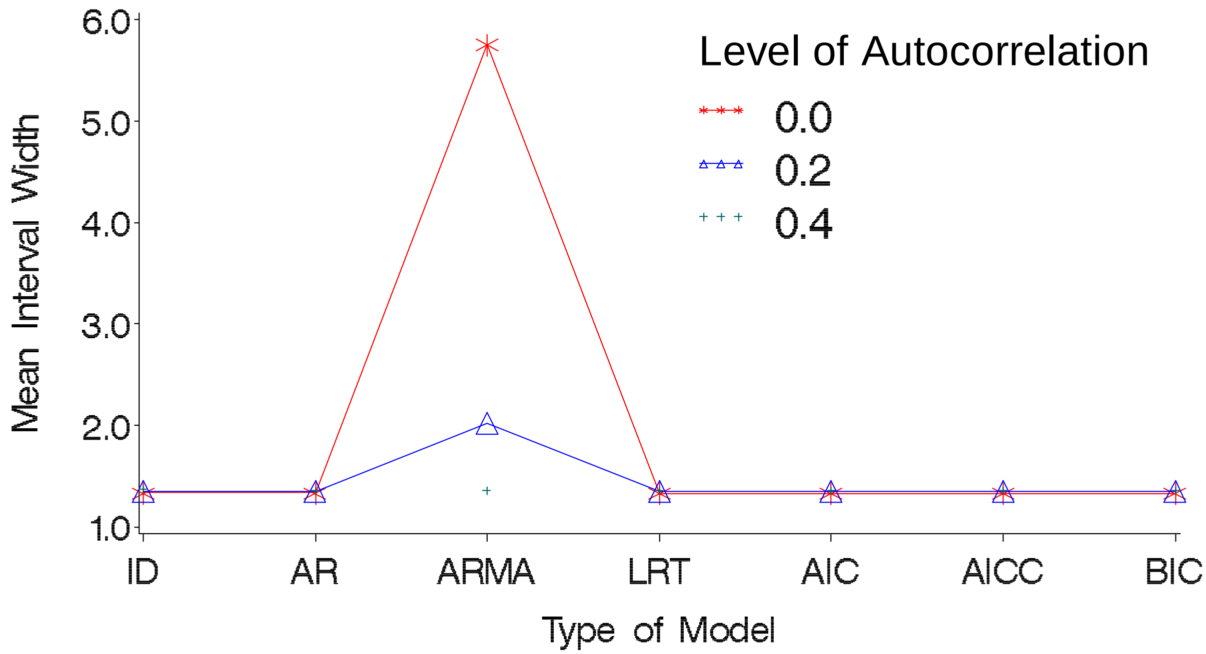

Figure 24. Line graph depicting mean confidence interval width as a function of the interaction effect between the type of model and the level of the autocorrelation parameter.

Lastly, the relationship for the interval width as a function of both the type of model and level of autocorrelation parameter was examined using a line graph. Figure 24 above illustrates that the confidence interval width is comparable across six of the seven models. However, the first order autoregressive moving average, ARMA (1, 1), model seemed to be an anomaly. Specifically, the mean confidence interval width becomes smaller as the level of the autocorrelation parameter increased, for $\rho = 0.0$ ($M = 5.76$, $SD = 4.97$), $\rho = 0.2$ ($M = 2.02$, $SD = 3.34$), and $\rho = 0.4$ ($M = 1.36$, $SD = 0.59$).

Overall average treatment effect for the interaction (shift in slopes). The average confidence interval width for the shift in slopes across the seven models was similar, except again for the first order autoregressive moving average model. The mean confidence interval width was 0.44 ($SD = 0.19$) for the six models; however, the mean was larger for the ARMA

(1,1) model (M = 0.81, SD = 1.20). Furthermore, the box plots (see Figure 25 below) reveal that there may be some variability within each model, particularly for the ARMA(1,1) model.

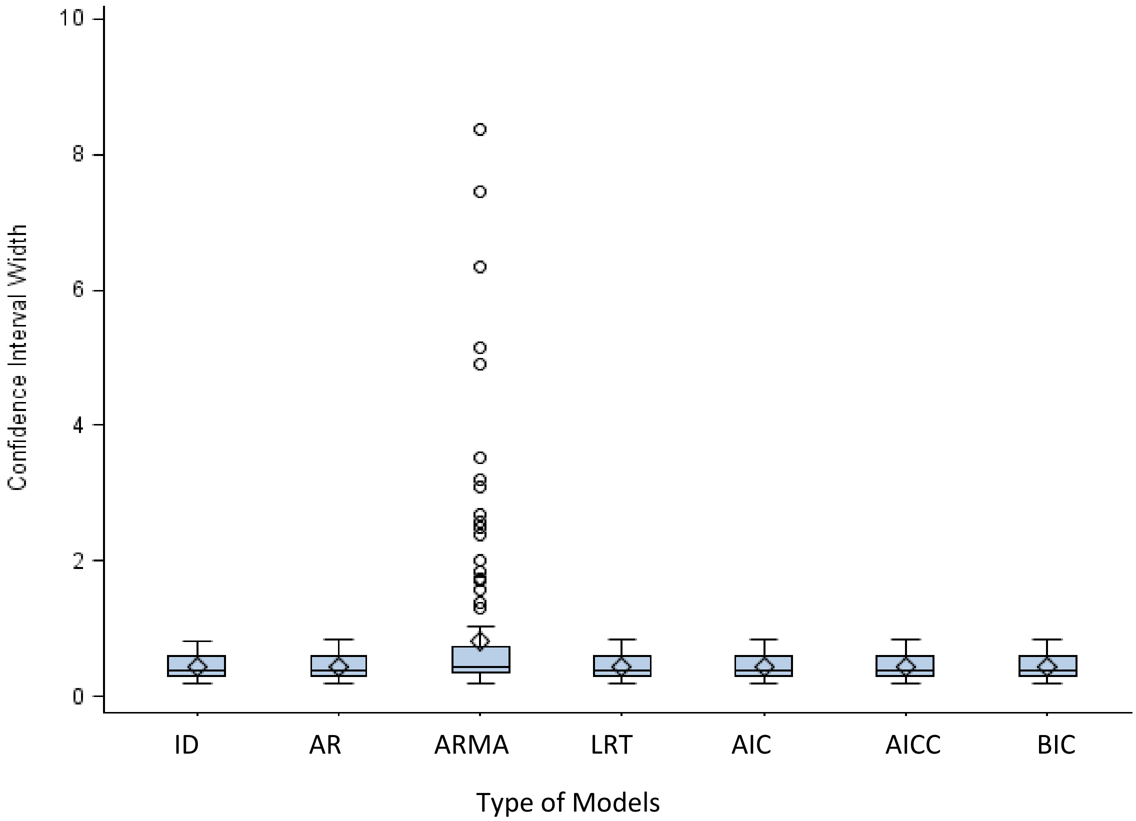

Figure 25. Box plot illustrating the distribution for the confidence interval width for the interaction effect (shift in slopes) across the seven models.

To further explore the variability in the mean confidence interval width for the interaction effect, GLM models were run. The results of the model, including 5-way interactions, explained 95% of the variability and revealed that there were three effects that constituted medium effects. These effects were as follows: variances of the error terms (η^2= .09), the number of primary studies included in the meta-analysis (η^2= .10), and the interaction of level of autocorrelation parameter and the type of model (η^2= .11). The means for the confidence interval width as a function of each of these effects are displayed in the Figures 26, 27, and 28 respectively.

The relationship for the variances of the error terms and the mean confidence interval width is illustrated in Figure 26 below. This relationship appears to be more direct, moreover, as the variances of the error terms shifted from most of the variance being at level-one (*M* = 0.34, *SD* = 0.26) to most of the variance appearing at the upper levels (*M* = 0.64, *SD* = 0.62), the mean confidence interval width increased.

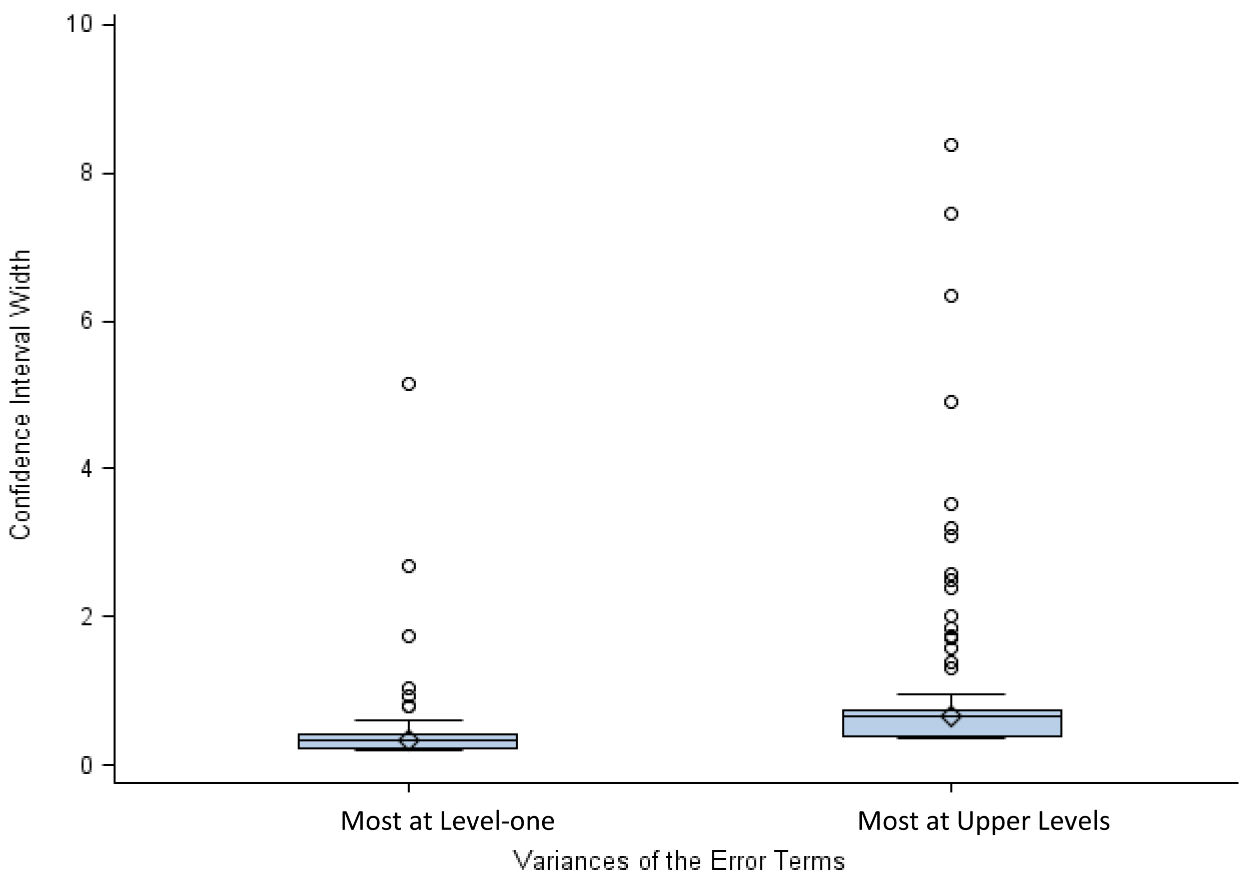

Figure 26. Box plot depicting relationship for confidence interval width for the shift in slopes as a function of the variances for the error terms.

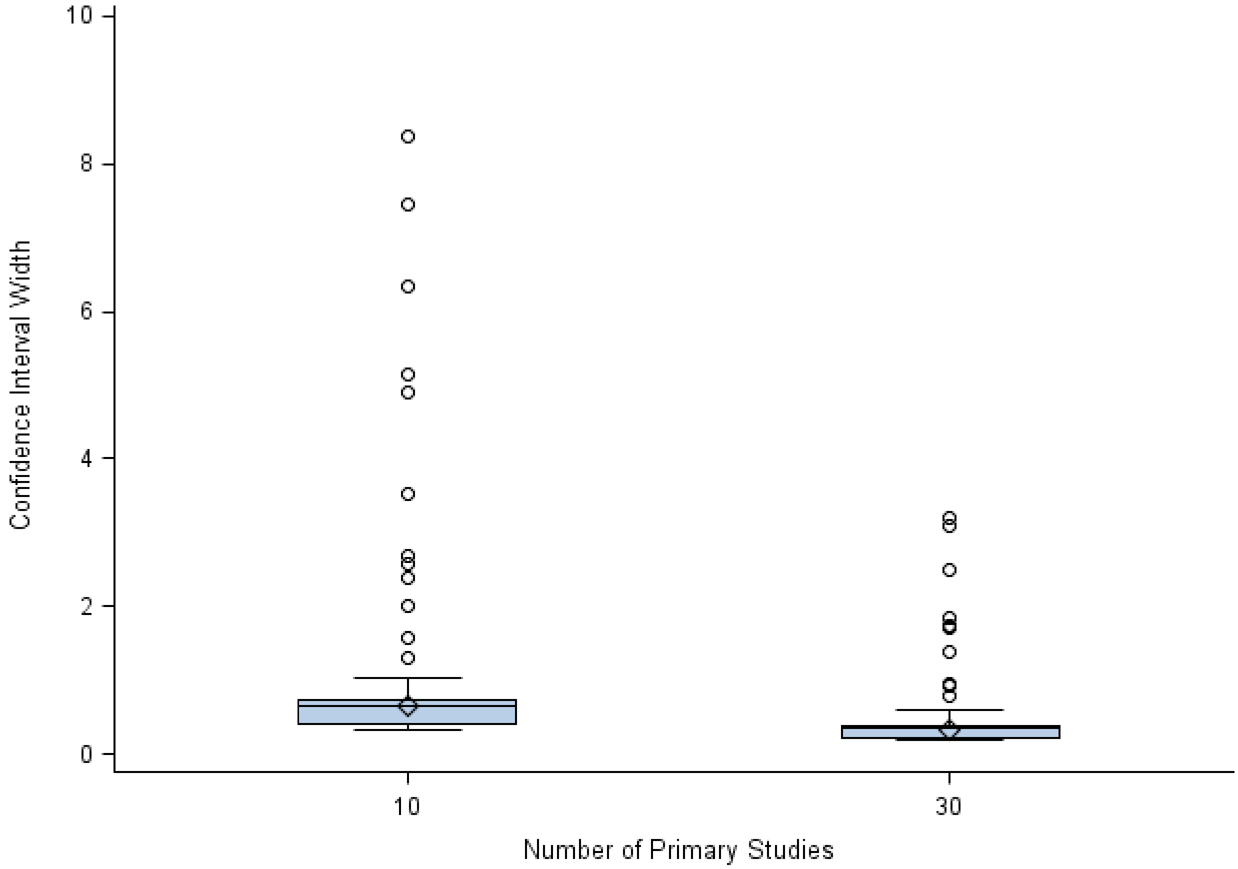

Figure 27. Box plot illustrating mean confidence interval width for the interaction effect as a function of the number of primary studies included in meta-analysis.

The mean confidence interval width as a function of the number of primary studies included in the meta-analysis is depicted in Figure 27 above. The figure shows that there is an inverse relationship, that as the number of primary studies increased then the mean confidence interval width decreased. Specifically, as the number of primary studies included in the meta-analysis increased from 10 to 30, then the mean confidence interval width decreased from 0.65 (*SD* = 0.62) to 0.34 (*SD* = 0.25).

Finally, the mean confidence interval width is displayed as a function of the interaction effect between type of model and the level of the autocorrelation parameter (see Figure 28 below). Similarly, the mean confidence interval width for the shift in slopes appeared comparable across all of the models, except for the first-order moving average parameter, ARMA (1, 1). According to the graph, the mean confidence interval width for the ARMA (1,1) model tended to decrease as the level for the autocorrelation parameter increased. Specifically, for ρ = 0.0 (*M* = 1.71, *SD* = 1.55), ρ = 0.2 (*M* = 0.71, *SD* = 1.35), and ρ = 0.4 (*M* = 0.45 , *SD* = 0.20).

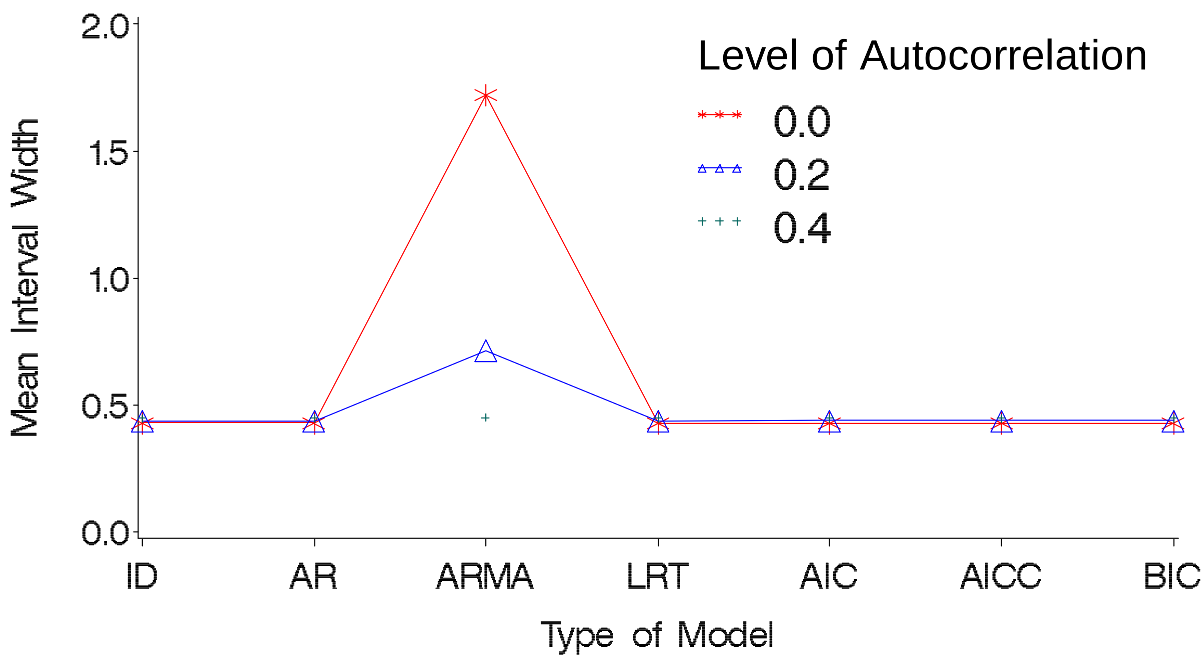

Figure 28. Line graph illustrates the mean confidence interval width for the shifts in slope as a function of the interaction effect between type of model and level of autocorrelation parameter.

Type I Error

The box plot depicting the distribution of the Type I error rates for the two intervention effects (shift in level and shift in slopes) across the seven models are displayed below in Figures 29 and 30, respectively.

Overall average treatment effect for phase (shift in level). The average type I error for the phase effect across the seven models was similar, the smallest mean Type I error was for the ARMA model (*M* = 0.049, *SD* = 0.003), while the largest mean Type I error was for the LRT model (*M* = 0.050, *SD* = 0.003). The means were comparable across models, with the type of model ($\eta^2 = 0.008236$), explaining very little of the total variability. Additionally, the range, 0.04 to 0.06 for the Type I error falls within Bradley's (1978) criterion. Therefore, no further analyses were warranted.

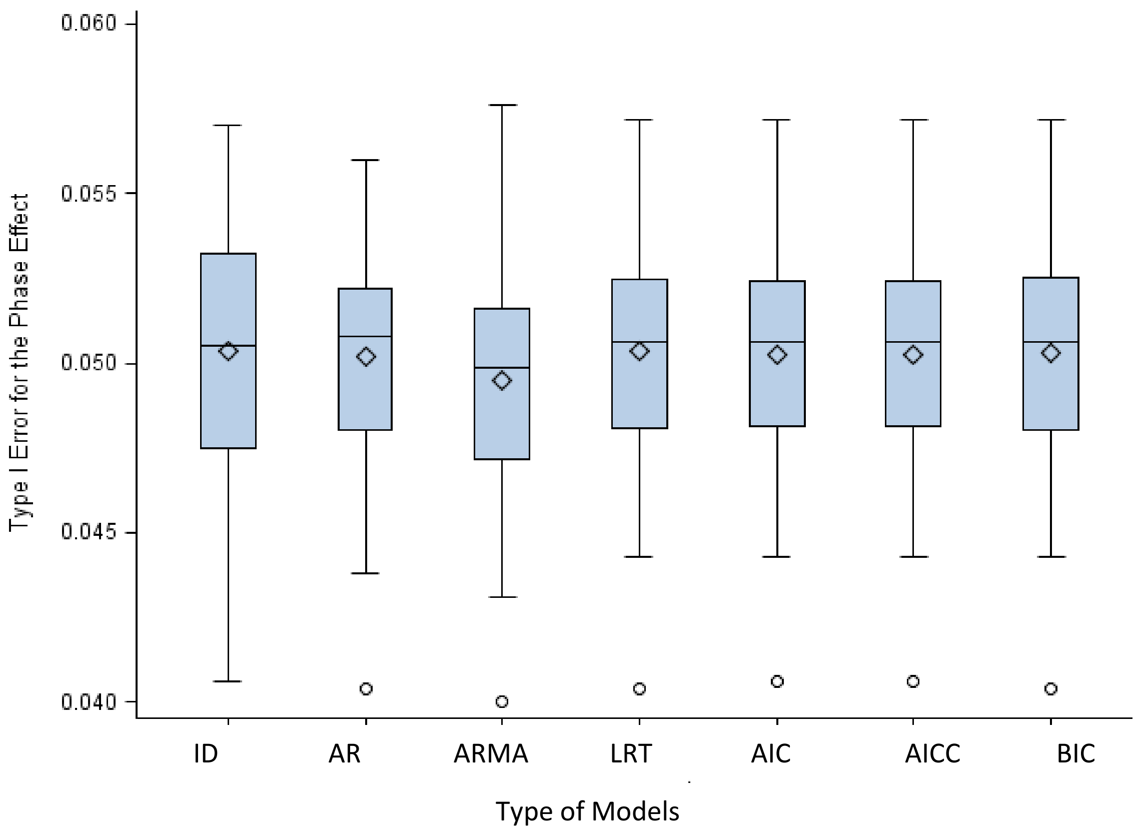

Figure 29. The distribution for the Type I error for the phase effect (shift in level) across the seven models.

Overall average treatment effect for the interaction effect (shift in slopes). The distribution for the Type I Error for the interaction effect is displayed in Figure 30 below. The average type I error for the interaction effect across the seven models was comparable, the largest mean Type I error for the LRT model (*M* = .050, *SD* = .004), conversely the smallest mean Type I error for the ARMA(1,1) model (*M* = .049, *SD* = .004). Again, the means were comparable across models, with the type of model ($\eta^2 = 0.027$), explaining very little of the total variability. Additionally, the range [0.036, 0.059] for the Type I error falls within Bradley's (1978) criterion; therefore, no further exploration was necessary.

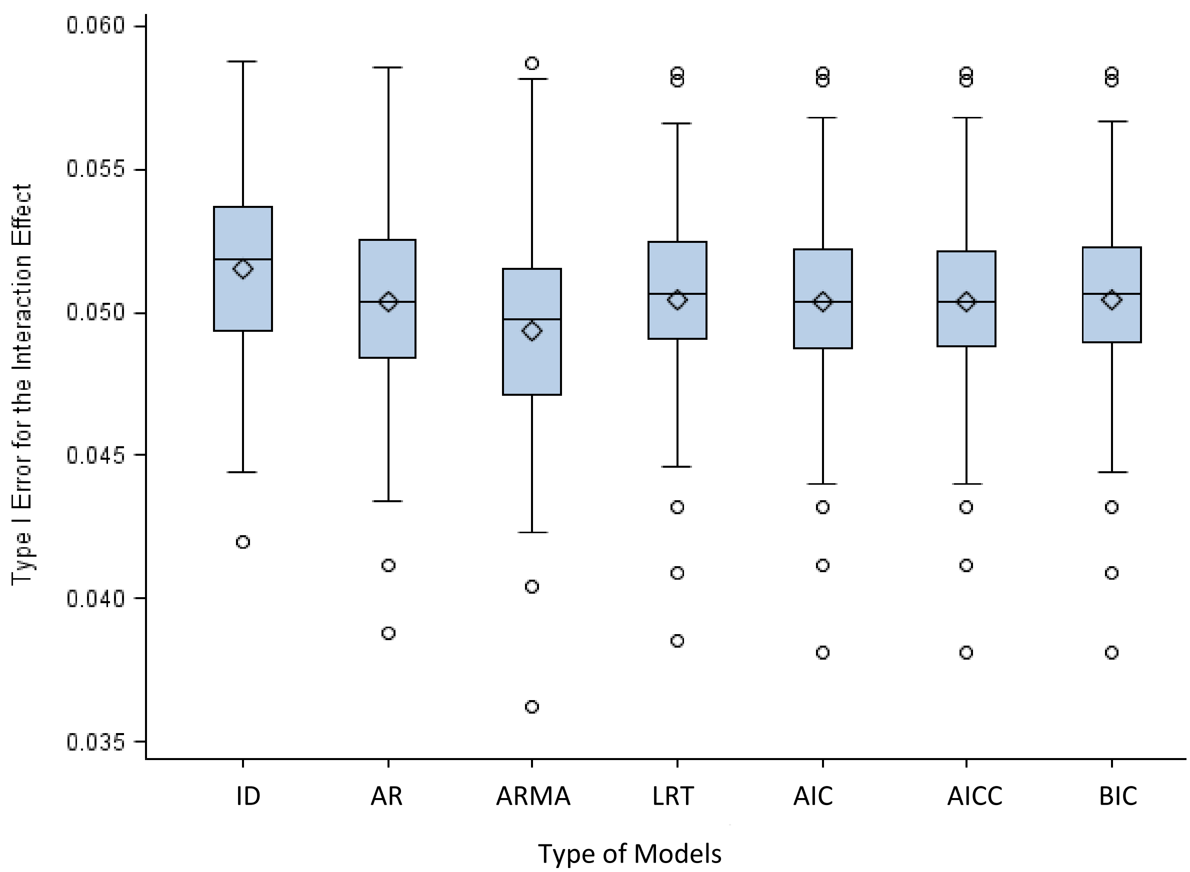

Figure 30. Box plots showing the distribution of the Type I error for the interaction effect (shift in slopes) for across the seven models.

Power for the Test of Fixed Effects

The box plot depicting the distribution of the power estimates for the two intervention effects (shift in level and shift in slopes) across the seven models are displayed below in Figures 31 and 33 respectively.

Overall average treatment effect for phase (shift in level). The distribution for the power estimates for the phase effect across the seven models was comparable. The smallest mean power estimates was for the ARMA model (*M* = 0.982, *SD* = 0.032), conversely, the largest mean power estimate was for the AR model (*M* = 0.983, *SD* = 0.031). The eta-squared (η^2 = .00015) for the type of model supports the small amount of variability among the seven models.

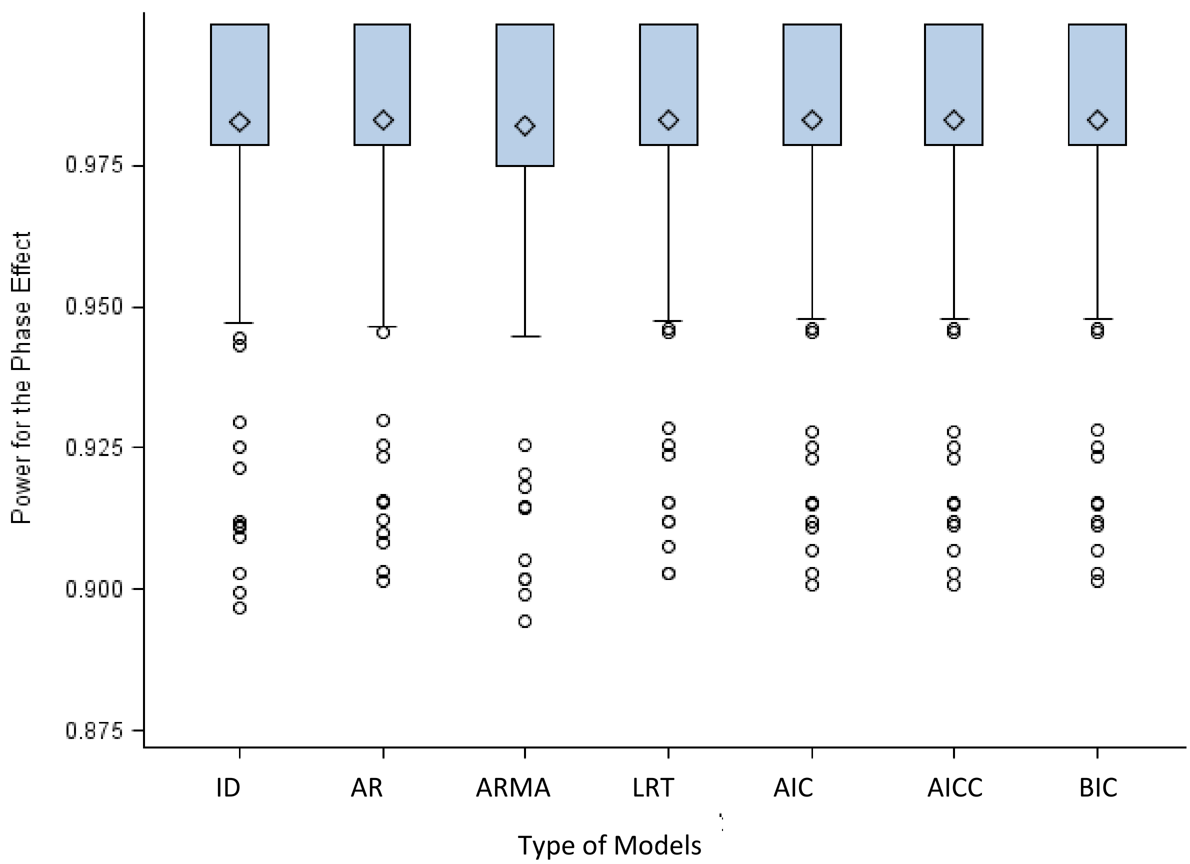

Figure 31. Box plots displaying the distribution for the power estimates for the phase effect across the seven models.

To further explore the variability, GLM models, including two-way interactions were used, this model explained 97% of the total variability. Additionally, the model revealed that the interaction effect between the number of primary studies and the variance of the error terms met the aforementioned criteria for being a medium effect (η^2= .30). Line graphs were then used to analyze the relationship between this effect and the mean power estimates.

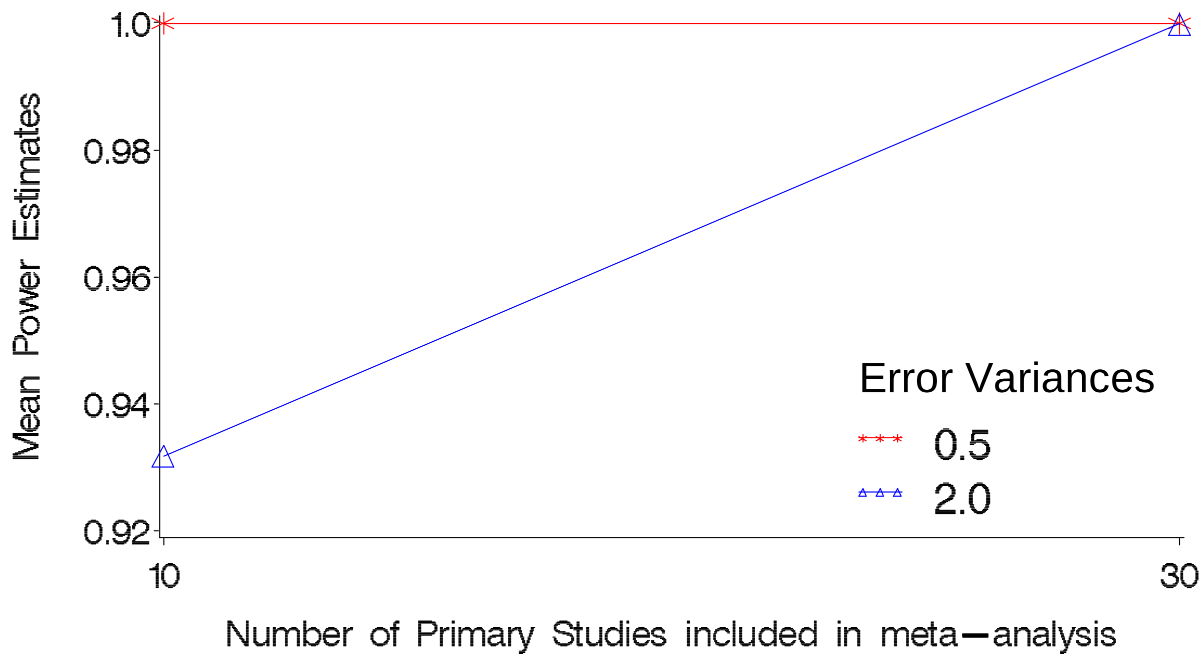

Figure 32. Line graphs illustrating the relationship between the mean power estimates and the interaction effect of the variances of the error terms and the number of primary studies included in the meta-analysis.

The line graphs above (see Figure 32) depicts that the mean power estimate is dependent upon both factors. For the cases when the error variances are mostly at the level-one, the effect on the power estimates does not depend on the number of primary studies included in the meta-analysis. However, when most of the error variances are at the upper levels, then when the number of primary studies shifts from 10 to 30 then the mean power estimates increase from *M* = 0.93, *SD* = 0.02 to *M* = 0.99, *SD* = 0.0008.

Overall average treatment effect for the interaction (shift in slopes). The average power estimate for the interaction effect (shift in slopes) across the seven models (η^2= .00002) was comparable.

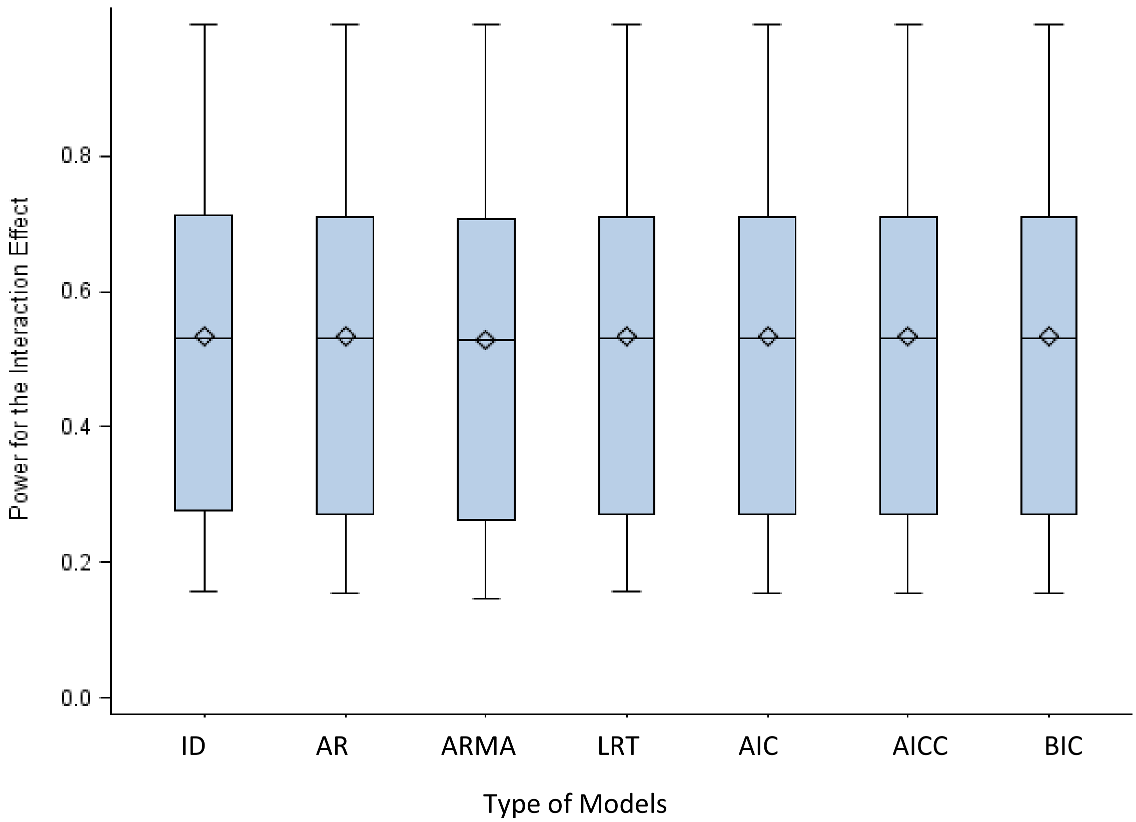

Figure 33. Box plots displaying the distribution for the power estimates for the interaction effect across the seven models.

To further examine the variability in the power estimates for the interaction effect, GLM models were created. The main effects only model explained 96% of the total variability, revealing two factors that met the criteria for being considered at least a medium effect: the variances of the error terms (η^2 = .42) and the number of primary studies included in the meta-analysis (η^2 = .51). The power estimates as a function of each of these factors are presented in Figures 34 and 35 below.

The power estimates as a function of the error variances are displayed below (see Figure 34). The power estimates decrease as the variances of the error terms are shifted from most of the variance at level-one (*M* = 0.71 , *SD* = 0.24) to most of the variance appearing at the upper levels (*M* = 0.36 , *SD* = 0.17).

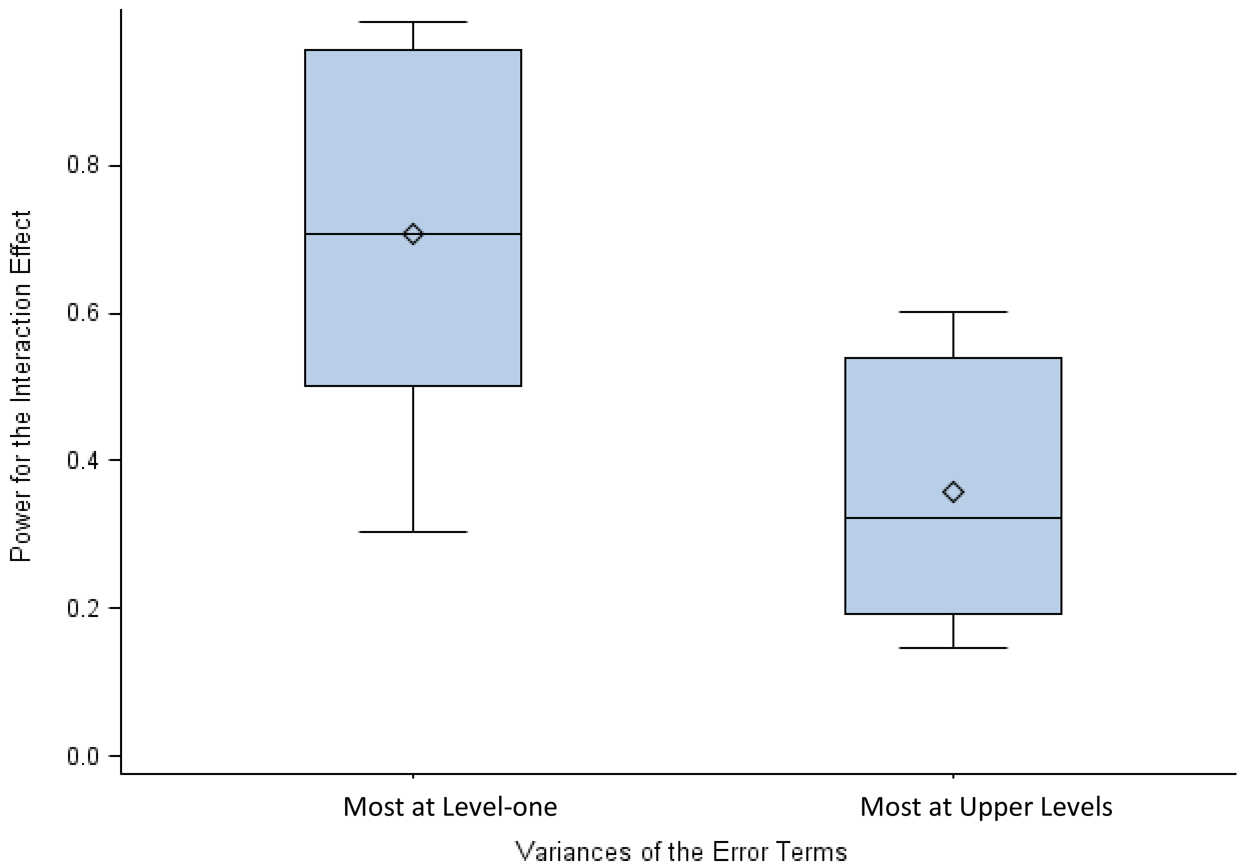

Figure 34. Box plots illustrating the relationship between the mean power estimates for the interaction effect and the variances of the error terms.

The power estimates as a function of the number of primary studies reveal that there is a direct relationship between the two parameters. As the number of primary studies increase from 10 to 30, then the power estimates increased from a mean of .34 (*SD* = 0.17) to a mean of .73 (*SD* = 0.21).

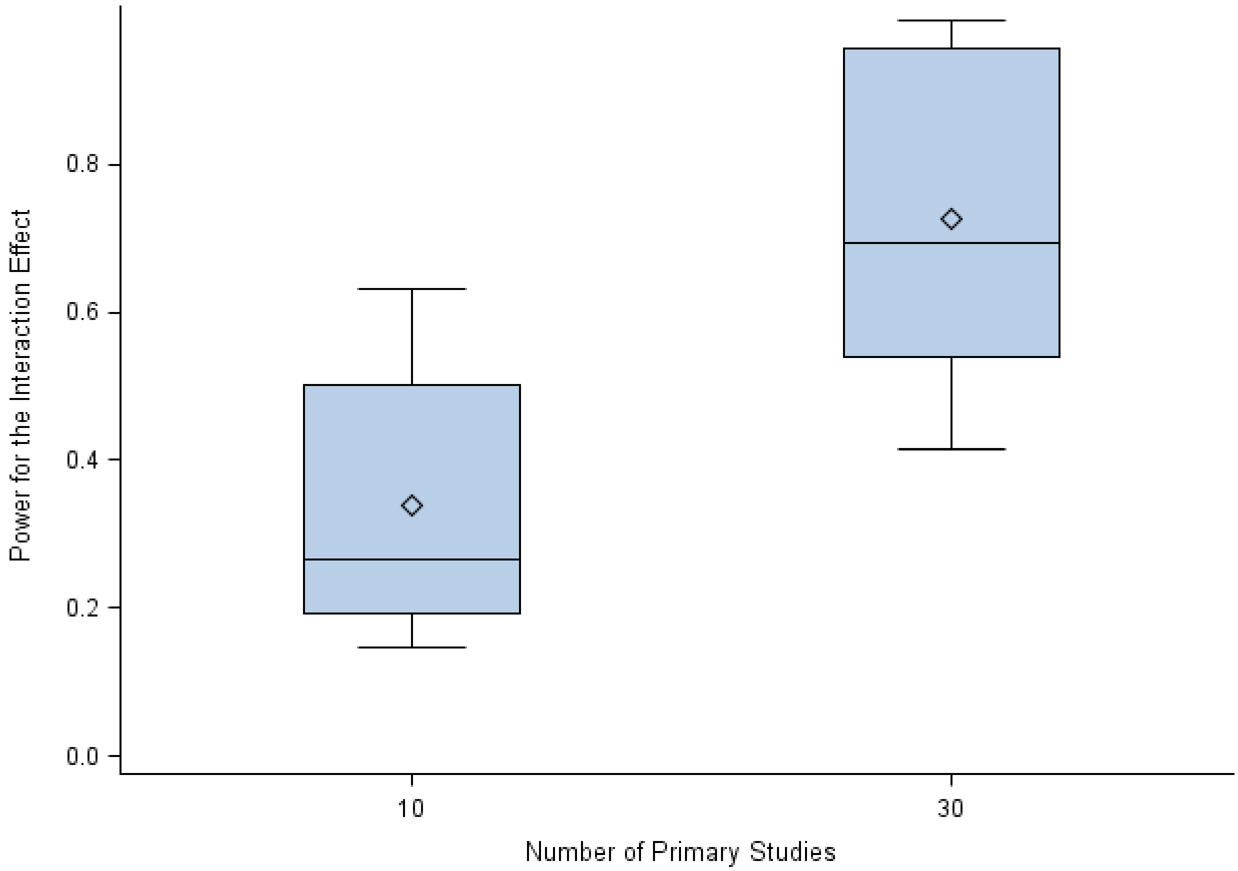

Figure 35. Box plots illustrating the relationship between the mean power estimates and the number of primary studies included in the meta-analysis.

Variance Components

Questions five and six were similar to questions two and three, except that the variance components were analyzed instead of the fixed effects. Specifically, question five examined the bias and RMSE associated with the variance components as a function of the seven factors used in this Monte Carlo study. The final question described the degree to which the confidence interval coverage and width for the variance components varied as a function of the seven design factors.

Relative Bias

The distribution of relative bias values for the variance components for the shift in level (phase) and the interaction effect (shift in slopes) is shown in Figure 36 and Figure 38 below. Relative bias was used for variables whose values were other than 1.0, but did not have levels that included 0. This enabled comparisons to be made across the different values of that variable. For example the levels for the variances of the error terms were 0.5 and 2.0. This simply involved dividing the bias estimates by the parameter value.

Level-three variance for the overall average treatment effect for the phase (shift in level). The average relative bias values were close to 0 across all of the models (the ID model and the first-order autoregressive moving average model). For the remainder of the models, the mean bias for the level-three variance components was similar, the eta-squared for the type of model = .007, further indicating that there was little variability among the type of model. Specifically, the average bias was the smallest (M = 0.02, SD = 0.02) for the ID model and largest was for the first-order autoregressive moving average model (M = 3.33, SD = 19.73). To further examine the variability of the bias for the level-three phase variance, GLM models were run. The model, including 5-way interactions, were run to see if any of the design factors had a significant effect, but none was found. Although the model explained 95% of the variability, none of the effects met the aforementioned criteria for a medium effect. Due to this finding, no further exploration was appropriate.

Due to the large amount of variability found in the bias estimates, the data were trimmed to produce plots where the distribution of the bias in the level-three phase effect can be examined for the design factors. Figure 37 below shows the distribution of the trimmed data, further analysis was conducted to see if there were any medium or larger effects. The results of these additional analyses are explicated in Appendix A.

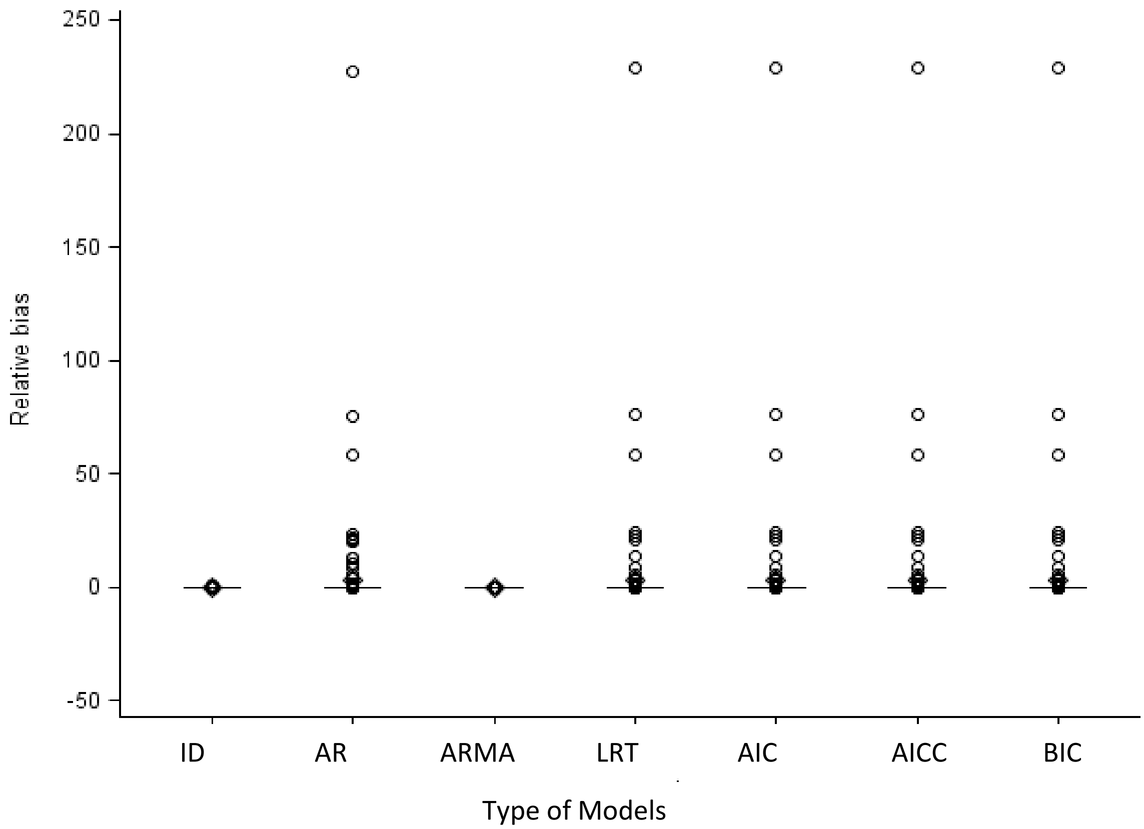

Figure 36. Box plots showing the distribution of the bias for the level-three shift in level (phase effect) across the seven models.

Level-three variance for the overall average treatment effect for the interaction effect (shift in slopes). The average relative bias values were close to 0 for the first-order autoregressive moving average model (*M* = 0.007, *SD* = 0.02). The relative bias mean estimate for the ID model was 0.32 (*SD* = 3.49). For the remainder of the models, the mean bias for the level-three variance components was comparable, the eta-squared for the type of model was .007, further indicating that there was little variability among the type of model.

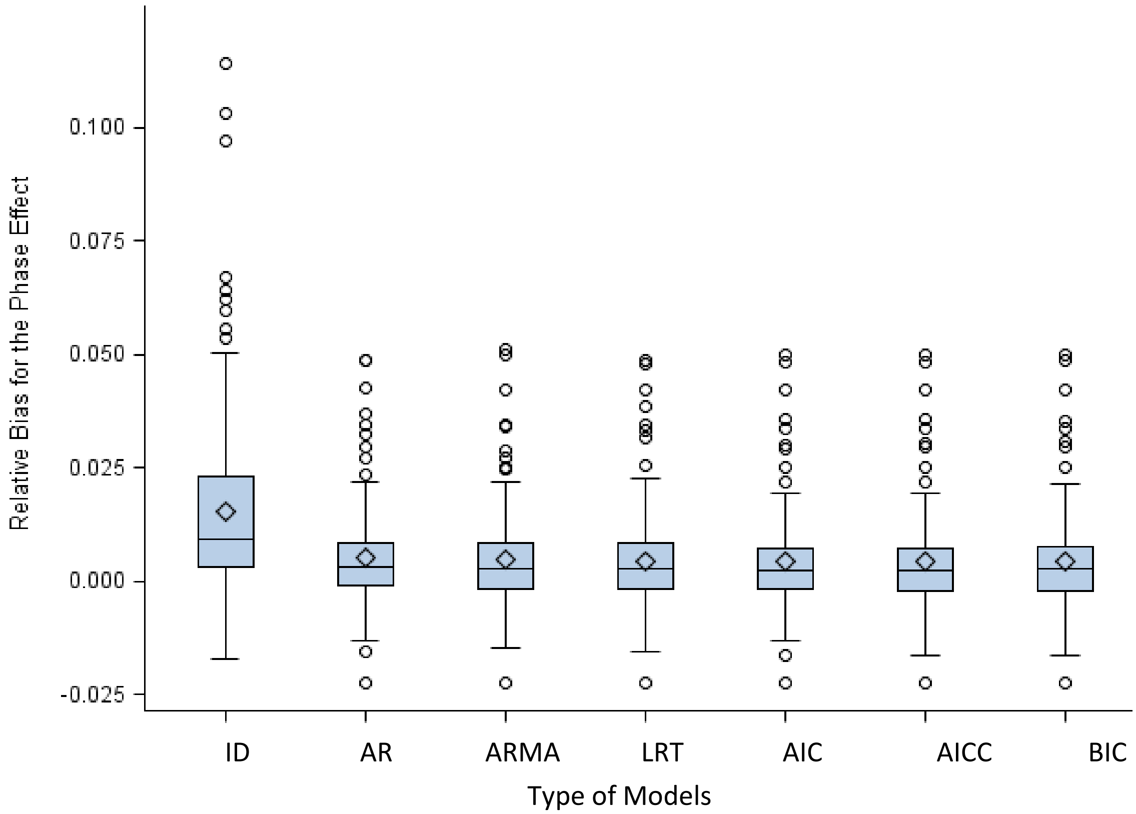

Figure 37. The box plot illustrating the trimmed distribution of the relative bias for the phase effect (shift in slopes) across the seven models.

To further examine the variability of the bias for the level-three phase effect, GLM models were run. The model, including 5-way interactions, explained 94% of the total variability, however none of the design factors had a medium effect, $\eta^2 > .0588$. Due to this finding, no further exploration was warranted on the original data. However, the data were trimmed for further analysis which is explained in detail below and in Appendix A.

Due to the large amount of variability found in the bias estimates, the data were trimmed to produce plots where the distribution of the bias in the level-three interaction effect can be examined for the design factors. Figure 39 below shows the distribution of the trimmed data for the level-three interaction effect for the shift in slopes, further analysis was conducted to see if there were any medium or larger effects on the bias by the study's design factors. The results of these additional analyses are thoroughly presented and explained in Appendix A .

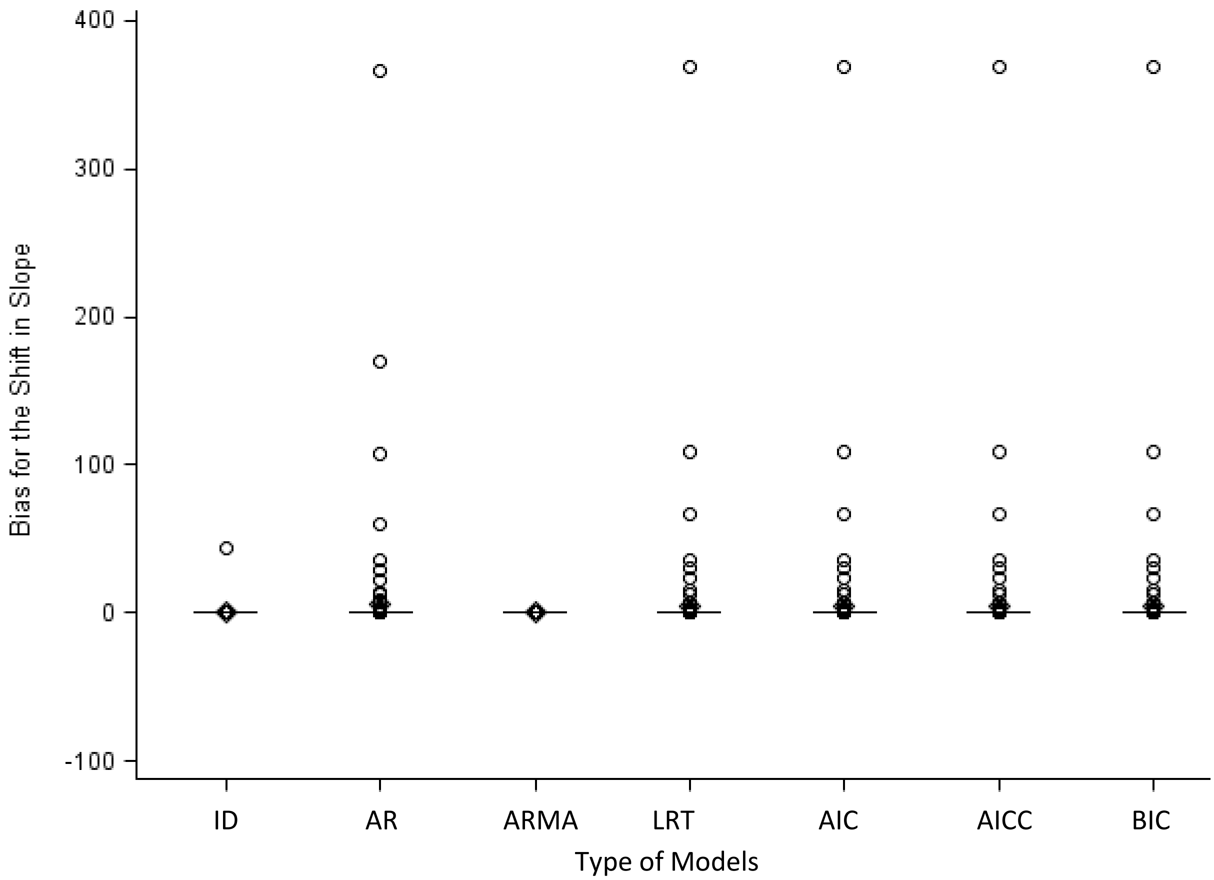

Figure 38. Box plots showing the distribution of the untrimmed bias for the level-three interaction effect (shift in slopes) across the seven models.

The box plots indicate that the means are similar across the seven models (η^2=.007), suggesting that the type of models did not explain a significant portion of the variability. GLM models were run to further examine the variability among the variance components. Furthermore, the model, including five-way interactions explained 94% of the total variability. However, none of the factors led to medium effects, indicating that the observed variability can be attributed to sampling error; and no further analyses was necessary.

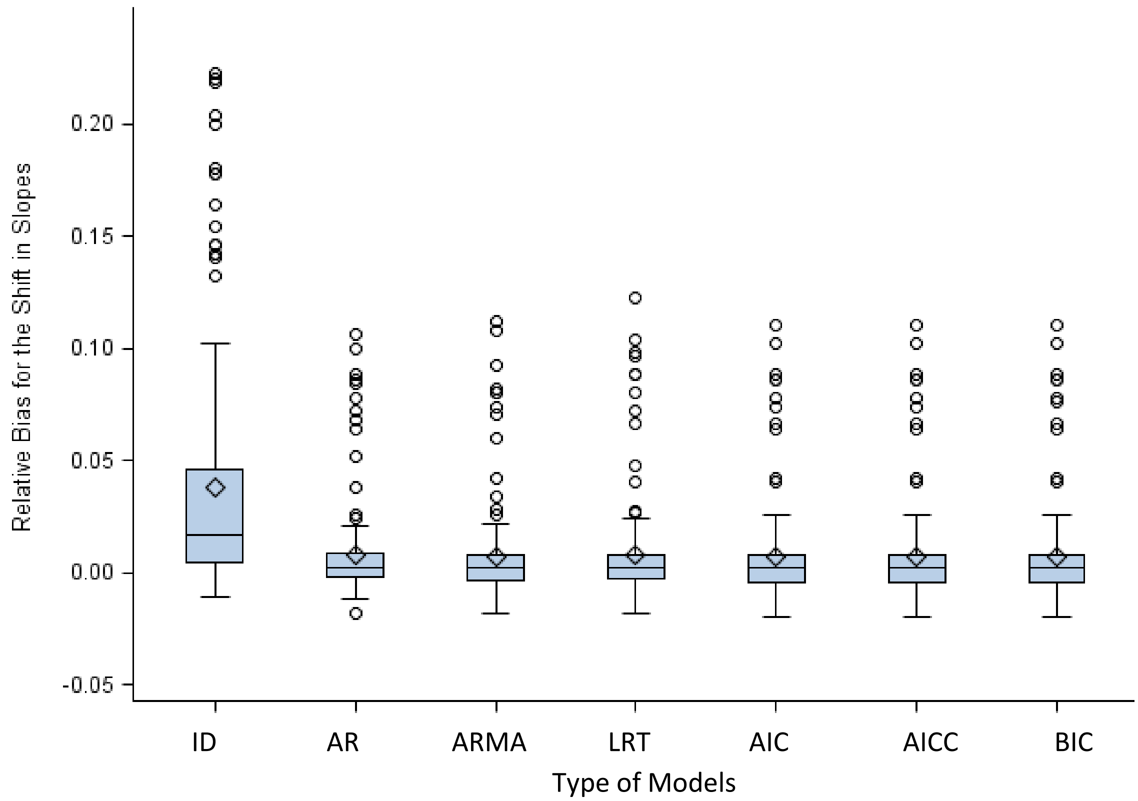

Figure 39. The distribution for the trimmed level-three variance for the overall average treatment effect for the interaction (shift in slopes).

Level-two variance for the average treatment effect for the phase (shift in level). The distribution for the level-two variance components for the overall average treatment effect for phase is displayed in Figure 40 below on page 123.

After trimming the data, which is displayed below in Figure 41, the box plots illustrate this distribution of the trimmed relative bias values for the level-two variance for the phase effect (shift in level) across the seven models. To further explore the effect of the design factors and the combination of the design factors, GLM models were run.

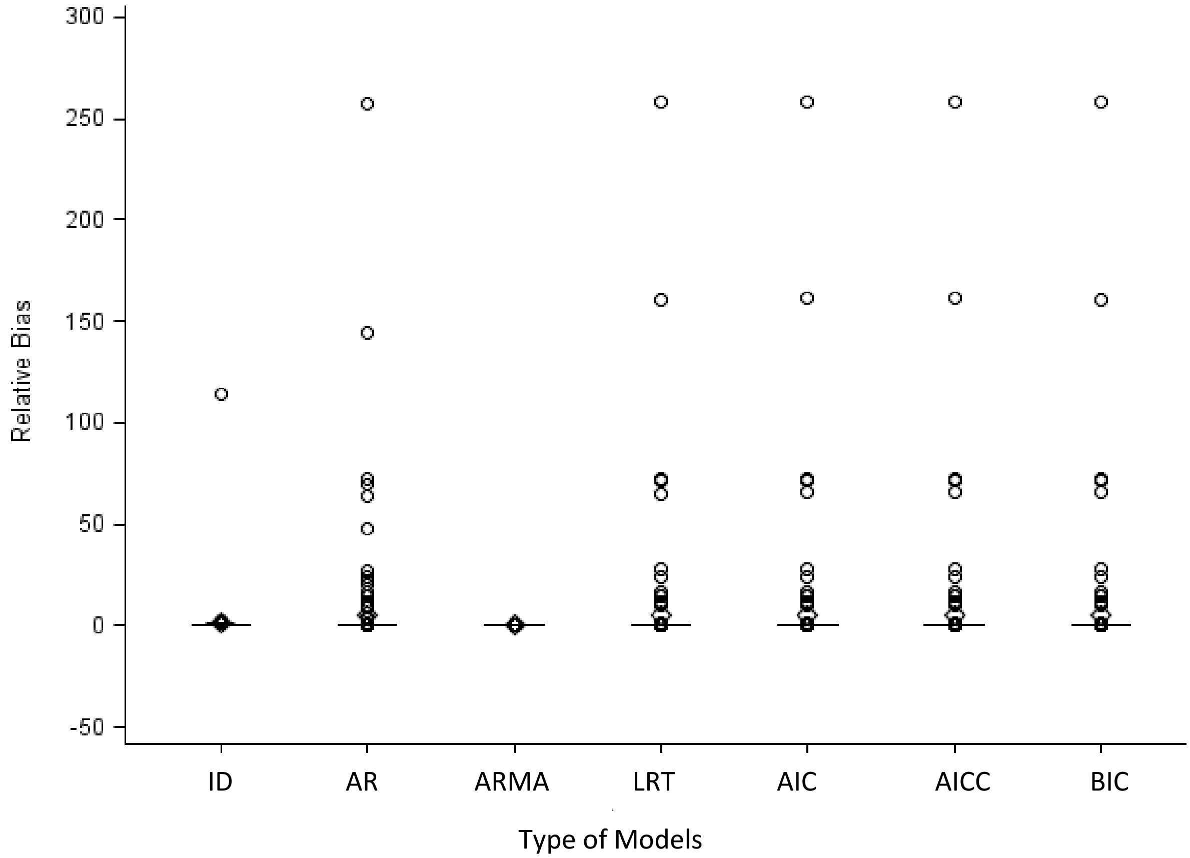

Figure 40. Box plot depicting the distribution for the level-two variance for the phase effect (shift in level) across the seven models.

The findings, (including the graphs) and explanations for the GLM model for this design factor is presented in greater detail in Appendix A. The results revealed one medium or larger effect: the interaction of the level of autocorrelation and the type of model (η^2= 0.22).

Line graphs were then produced to further investigate this relationship and found that the trimmed relative bias for the level two variance for the phase effect (shifts in level) was comparable across all of the models as the level of the autocorrelation increased However, for the ID model, the mean relative bias tended to increased greatly as the level of autocorrelation increased.

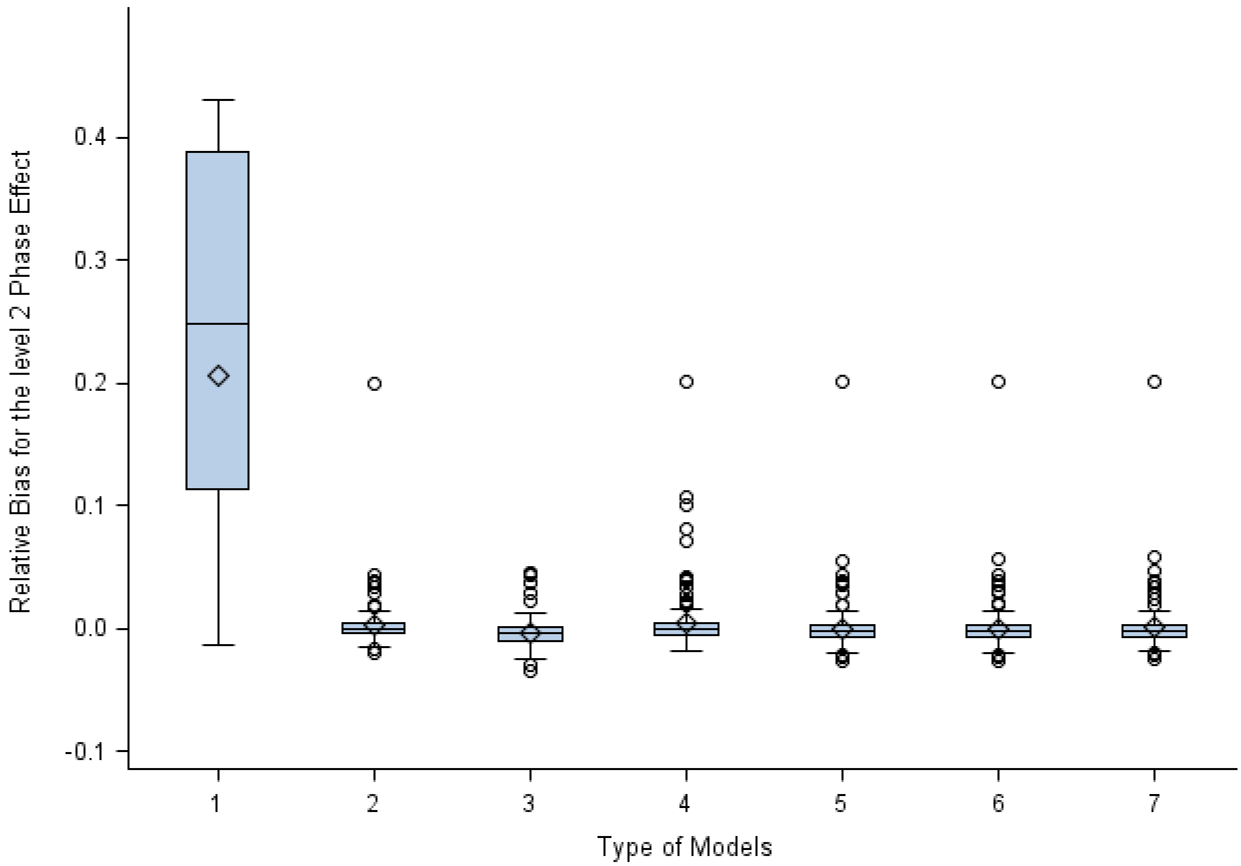

Figure 41. Box plot representing the trimmed distribution for the level-two phase effect (shift in slopes) bias across the seven models.

Level-two variance for the overall average treatment effect for the interaction (shift in slopes). The distribution illustrating the distribution for the bias for the level-two variance components for the interaction effect is displayed in Figure 42 below. The plots reveal that the means across the models (η^2 = .007) are similar. The largest mean for relative bias was observed for the first-order autoregressive model (*M* = 8.32, *SD* = 44.89) and the smallest mean was for the ID model, *M* = 0.64, *SD* = 3.45.

To further explore the variability, GLM models were run. The model, including five-way interactions explained 93% of the total variability and revealed no substantial effects. Therefore, no further analyses were necessary on the original data.

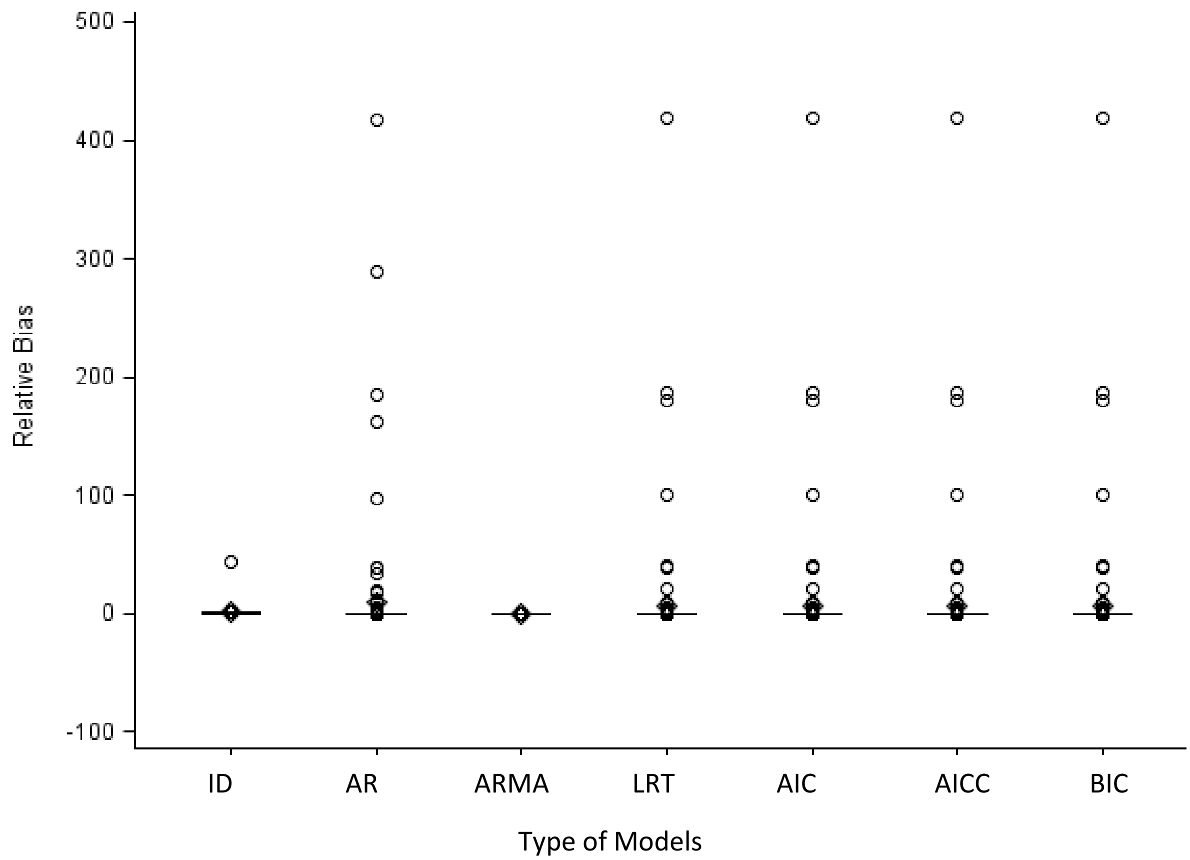

Figure 42. Box plots illustrating the distribution for the level-two variance component for the interaction effect (shift in slopes) across the seven models.

Due to the extreme variability noted in the graphs, the data were trimmed so that the relative bias could be further investigated. The box plots below (see Figure 43) show the distribution of the trimmed data. Further analyses were run to further examine the relationship with the trimmed data and the design factors (the results of the additional analyses can be found in Appendix A).

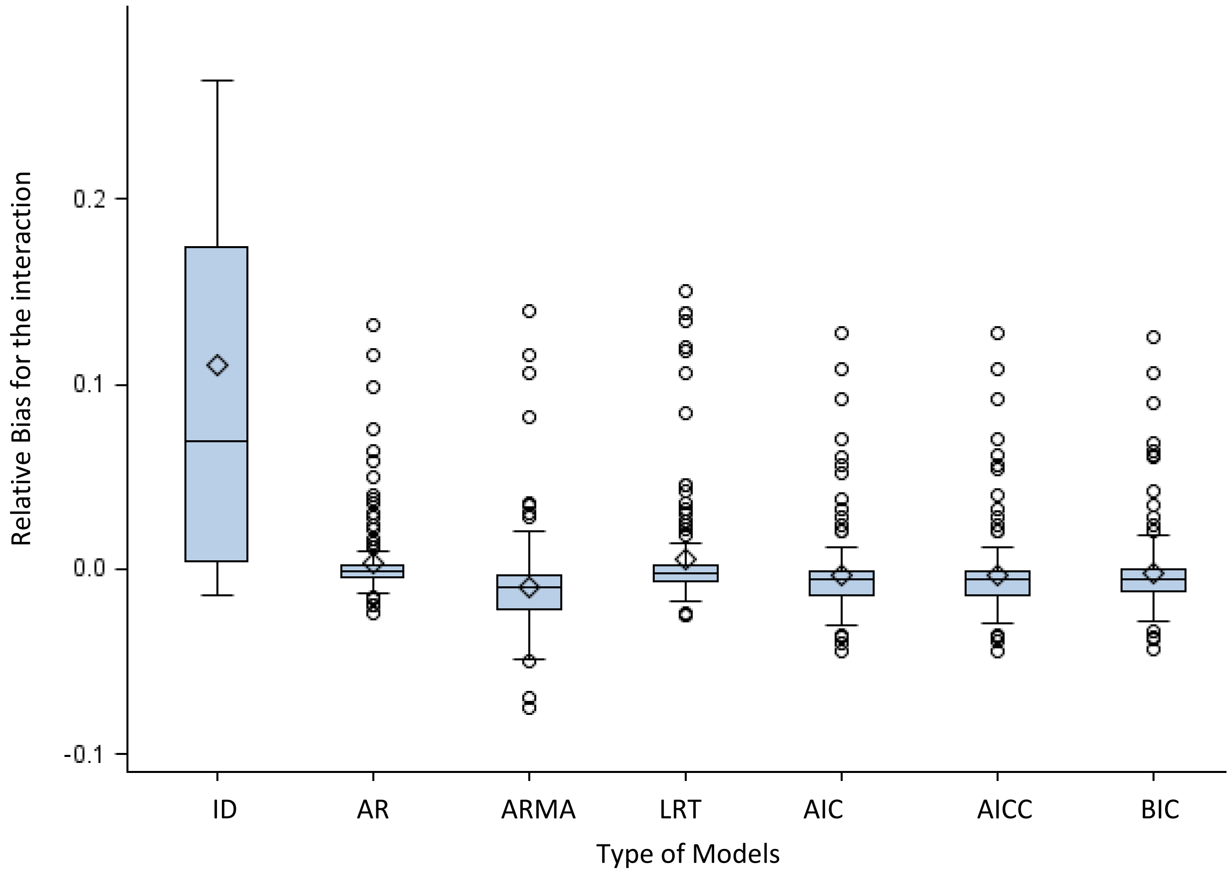

Figure 43. The distribution of the trimmed relative bias for the level-two variance for the interaction.

Level-one or Residual Variance. The distribution for the level-one or residual variance is displayed in Figure 44 below across the seven models. The plots revealed that there is variability across models in the mean bias for the level-one variance. The largest mean bias estimate is observed for the first-order autoregressive moving average model (*M* = 0.20, *SD* = 0.15), conversely, the smallest mean bias estimate is seen for the ID model (*M* = -0.08, *SD* = 0.06). To further explore the variability in the bias estimates, GLM models were run. The model, including 5-way interactions, explained 94% of the total variability, and revealed the following medium effect: the interaction between the level of the autocorrelation parameter and the type of model (η^2 = .08).

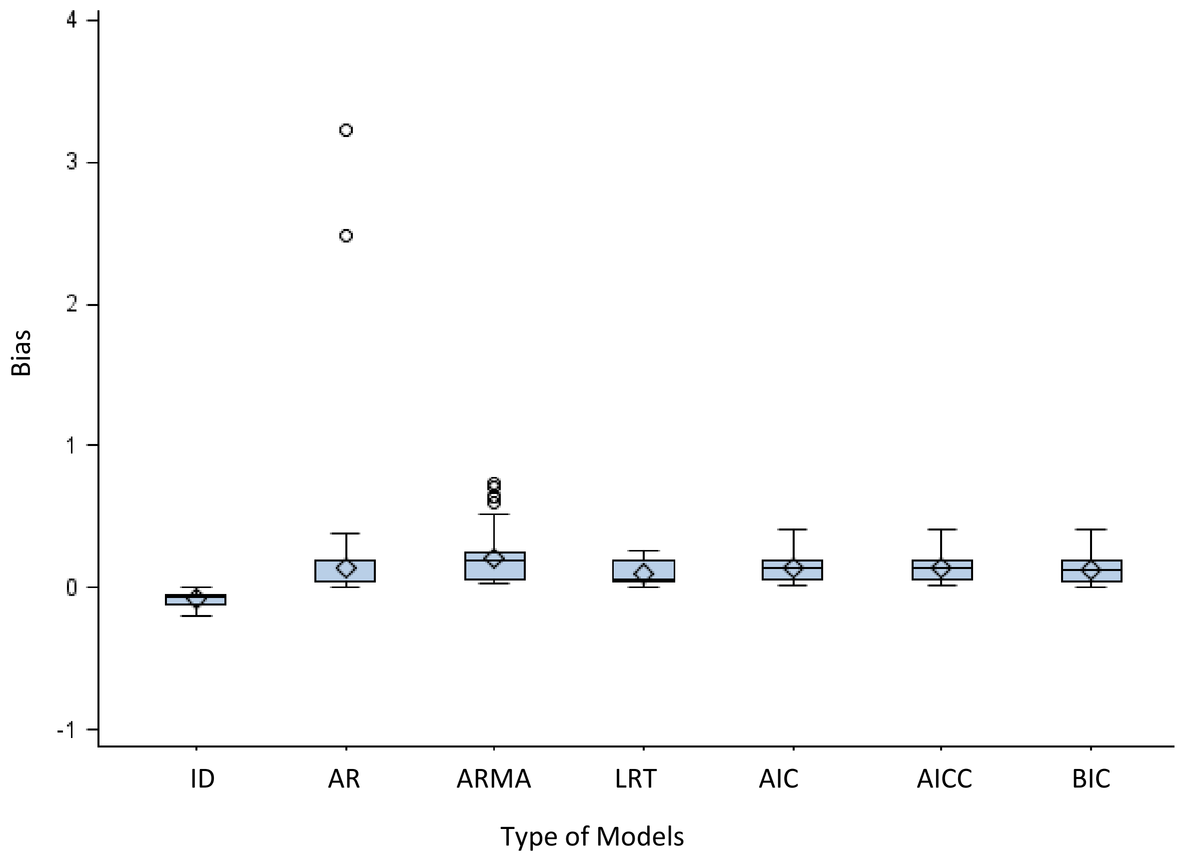

Figure 44. Box plots illustrating the distribution for the level-one variance (residual variance) across the seven models.

Line graphs were then created to examine the relationship between the bias in the level-one variance and the interaction effect between the level of autocorrelation and the type of model. Figure 45 below illustrates this relationship, depicting that the effect of the level of the autocorrelation parameter on the level-one bias estimates depends on the type of model. Moreover, as the level of the autocorrelation parameter increased, the bias increased for five of the models (the AR models and the models selected by each of the four fit indices). This correlation was similar for the ID model, but in the inverse direction. Specifically, as the level of the autocorrelation increased, then the bias increased for the ID model, but in the negative direction. Lastly, the ARMA model had the least bias for the level-one variance when ρ = 0.2, then the mean bias increased when ρ = 0.4, and then largest for when the autocorrelation parameter was 0.

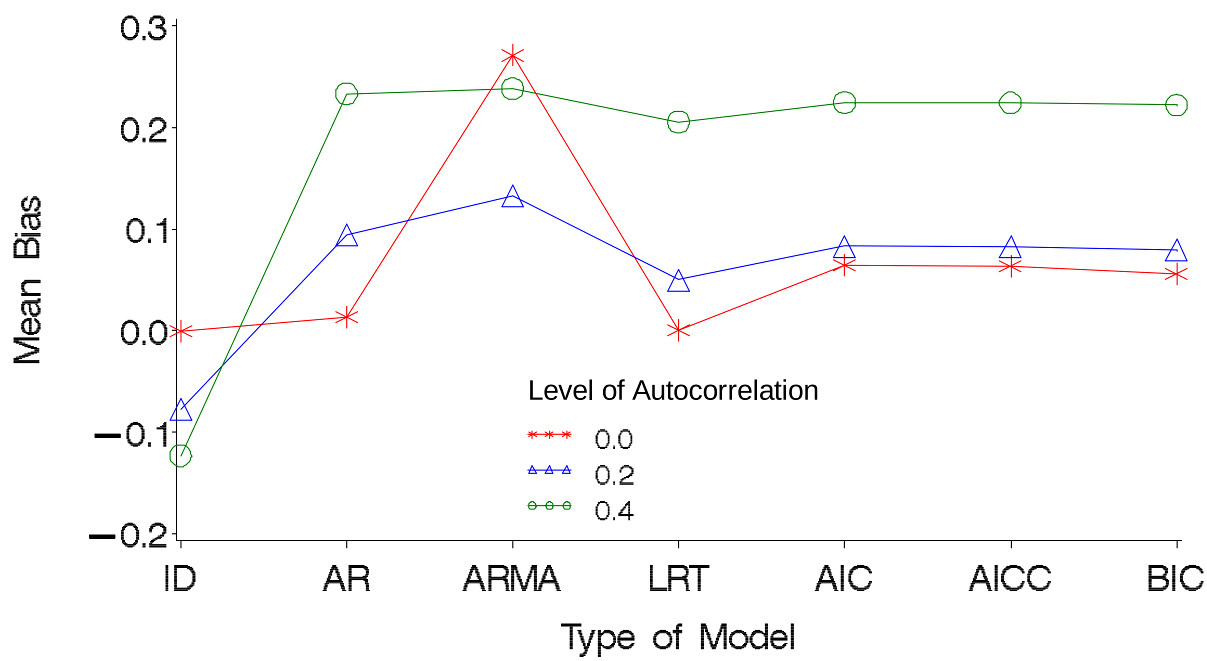

Figure 45. Line graph illustrating the mean bias in the level-one variance (residual variance) and the interaction effect of the level of the autocorrelation parameter and the type of model.

Autocorrelation Parameter. The box plot (see Figure 46 below) depicting the distribution of the bias for the autocorrelation parameter across all of the models, except for the ID model (where the autocorrelation parameter was estimated to be 0). As observed in the Figure 46, the means for the bias of the autocorrelation parameter varied across all models. The smallest mean bias for the autocorrelation parameter was observed for the AR model (*M* = -0.0007, *SD* =

0.002), the largest mean bias was observed for the first-order autoregressive moving average, ARMA (1,1), model (*M* = -0.03, *SD* = 0.08).

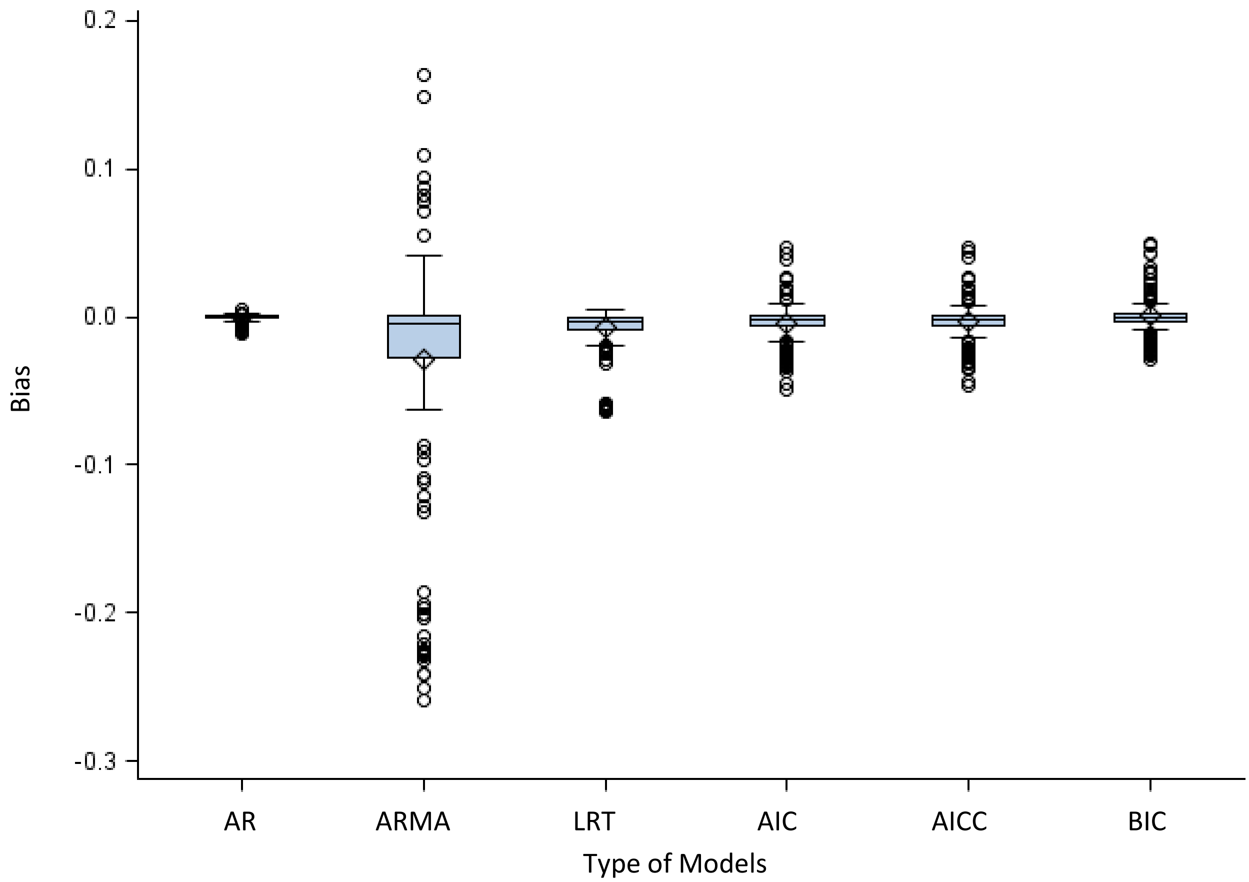

Figure 46. Box plot depicting the distribution of the bias for the autocorrelation parameter across the six models.

To further examine the variability, GLM models were run. The model, including three-way interactions explained 94% of the total variability. The results of the model indicated that the three-way interaction among the variances of the error terms, the level of the autocorrelation parameter, and the type of model ($\eta^2 = 0.12$) met the aforestated criteria for being a medium effect. Line graphs were then created to further explore the relationship of the means for the autocorrelation parameter across this interaction effect.

In Figure 47 below, the means for the bias of the autocorrelation parameter across the six models for when most of the variance is at the upper levels is displayed. The mean bias for the autocorrelation parameter was minimal across five of the models across all levels of the autocorrelation parameter. The variability in the mean bias was greater for the ARMA (1,1) model across the various levels of the autocorrelation parameter. For this model, the bias was minimal, *M* = -0.0004, *SD* = 0.006, when the autocorrelation parameter was 0.4. The bias then increased as the autocorrelation parameter increased from 0.0 (*M* = -0.007, *SD* = 0.002) to 0.2 (*M* = -0.01, *SD* = 0.02) for the ARMA (1,1) model. The relationship is very similar for when most of the variance is at level-one, however the mean bias is greater in the negative direction when the autocorrelation parameter is equal to 0.

First-Order Autoregressive Moving Average Parameter. The distribution of the bias for the moving average parameter is displayed in Figure 48 below. The means across the five models (the moving average parameter was estimated to be zero for both the ID and the AR models) varied. As illustrated by the figure, the mean bias for the moving average parameter is smallest for the models selected by the AIC and AICC (*M* = -0.067, *SD* = 0.154). The mean bias is largest for the first-order autoregressive moving average model, *M* = 0.16, *SD* = 0.15.

To further examine the variability in the bias of the moving average parameter, GLM models were run. The main-effects only model explained 94% of the total variability and revealed that there were two significant medium effects: the type of model ($\eta^2 = 0.27$) and the level of the moving average parameter ($\eta^2 = 0.61$). Graphs were then created to represent the relationship between the mean bias for the moving average parameter and each of the effects.

The relationship between the mean bias of the moving average parameter and the level of the moving average parameter is depicted below (see Figure 49). The graph shows that as the level of moving average parameter increased from 0.0 (*M* = 0.08, *SD* = 0.11) to 0.2 (*M* = -0.12, *SD* = 0.08) to 0.4 (*M* = -0.27, *SD* = 0.14), the mean bias also increased for the moving average parameter.

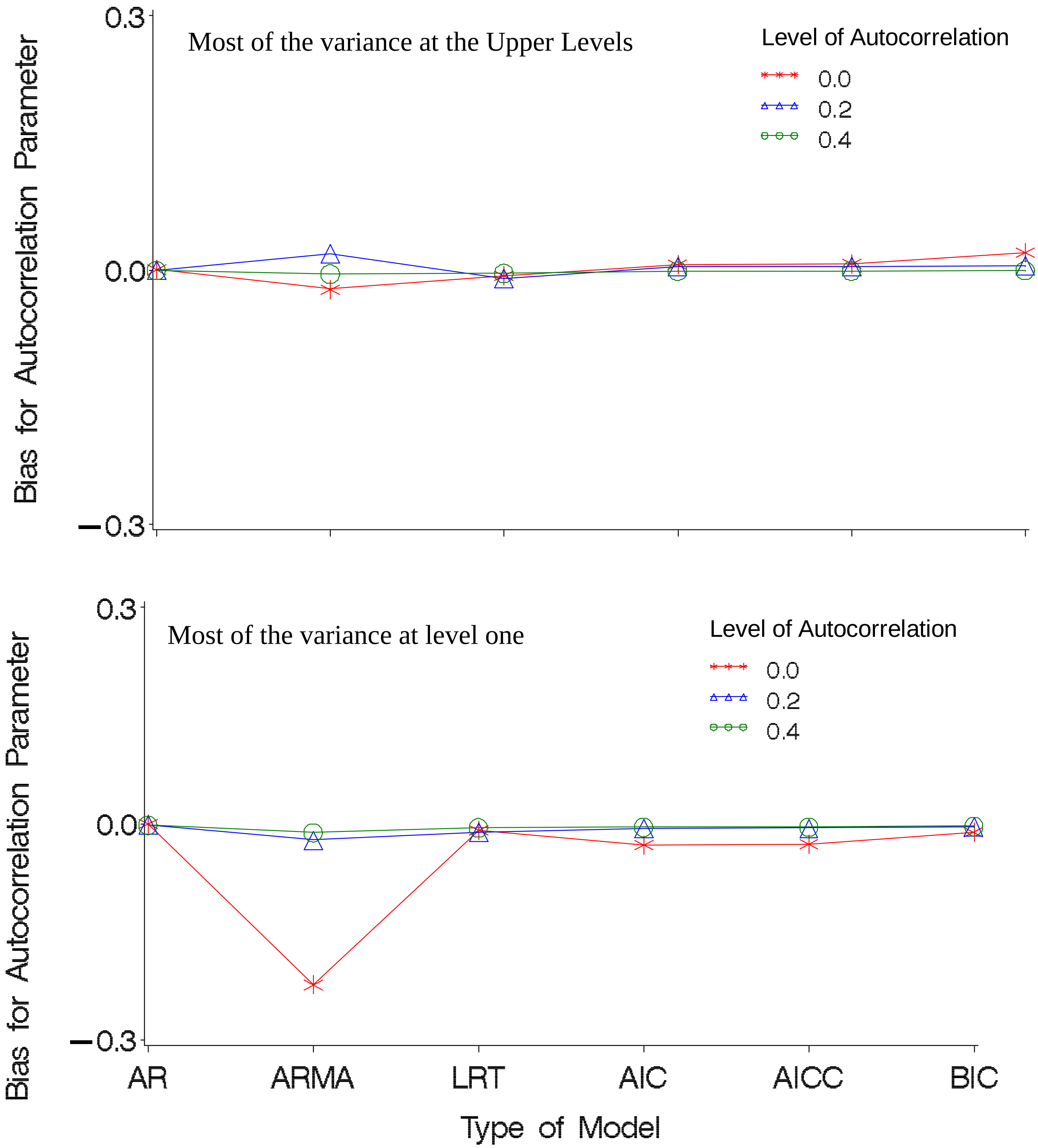

Figure 47. Line graphs illustrating the means for the relationship of the bias in the autocorrelation parameter and the three-way interaction among the variances of the error terms, the level of the autocorrelation parameter, and the type of models.

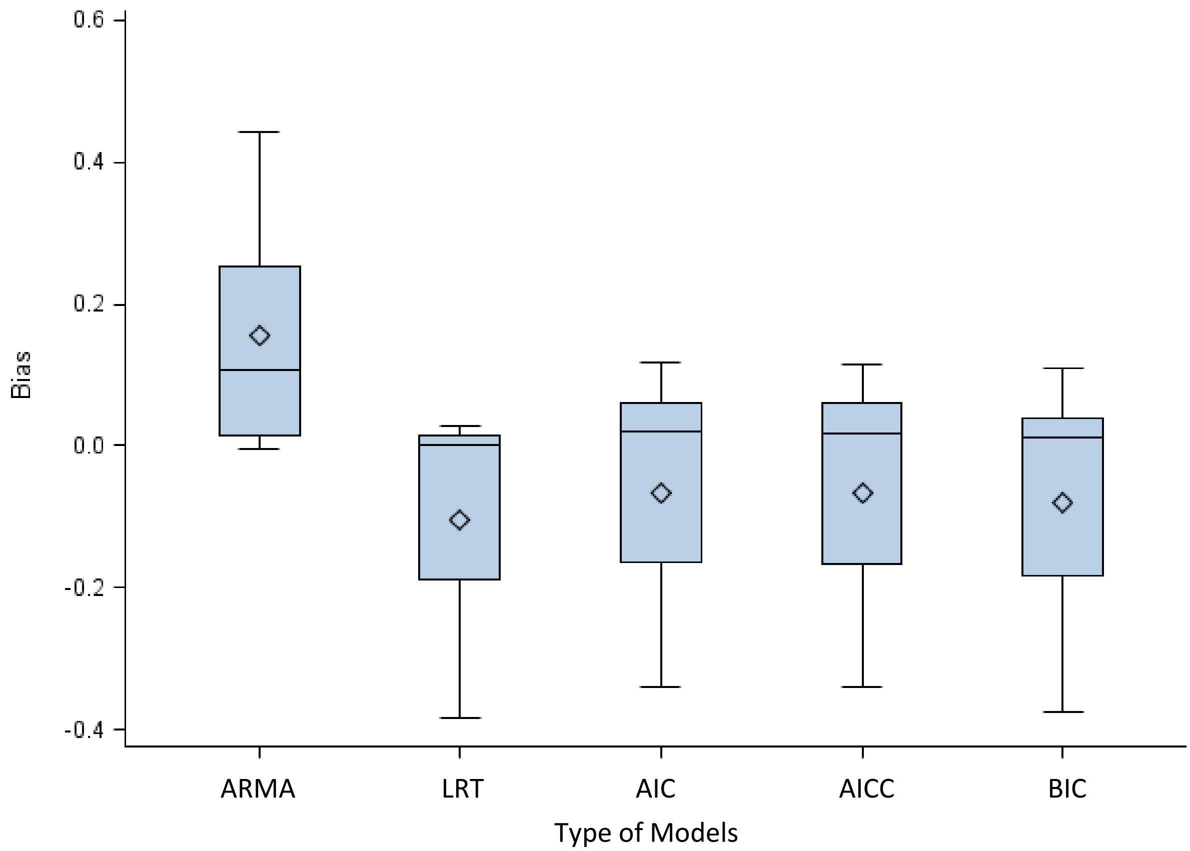

Figure 48. Box plots showing the distribution for the bias in moving average parameter estimate across the five models.

Root Mean Square Error (RMSE)

The distribution of the RMSE values for the variance components for the shift in level (phase) and the interaction effect (shift in slopes) is shown in Figure 50 and Figure 52 respectively.

Level-three variance for the overall average treatment effect for the phase (shift in level). The box plot in Figure 50 below illustrates the distribution of the RMSE values for the level-three variance for the phase effect across the seven models. The smallest mean RMSE value was observed for the ARMA(1,1) model (*M* = 0.64, *SD* = 0.40), however the largest mean RMSE value was noted for the AR(1) model (*M* = 382.79, *SD* = 3098.68).

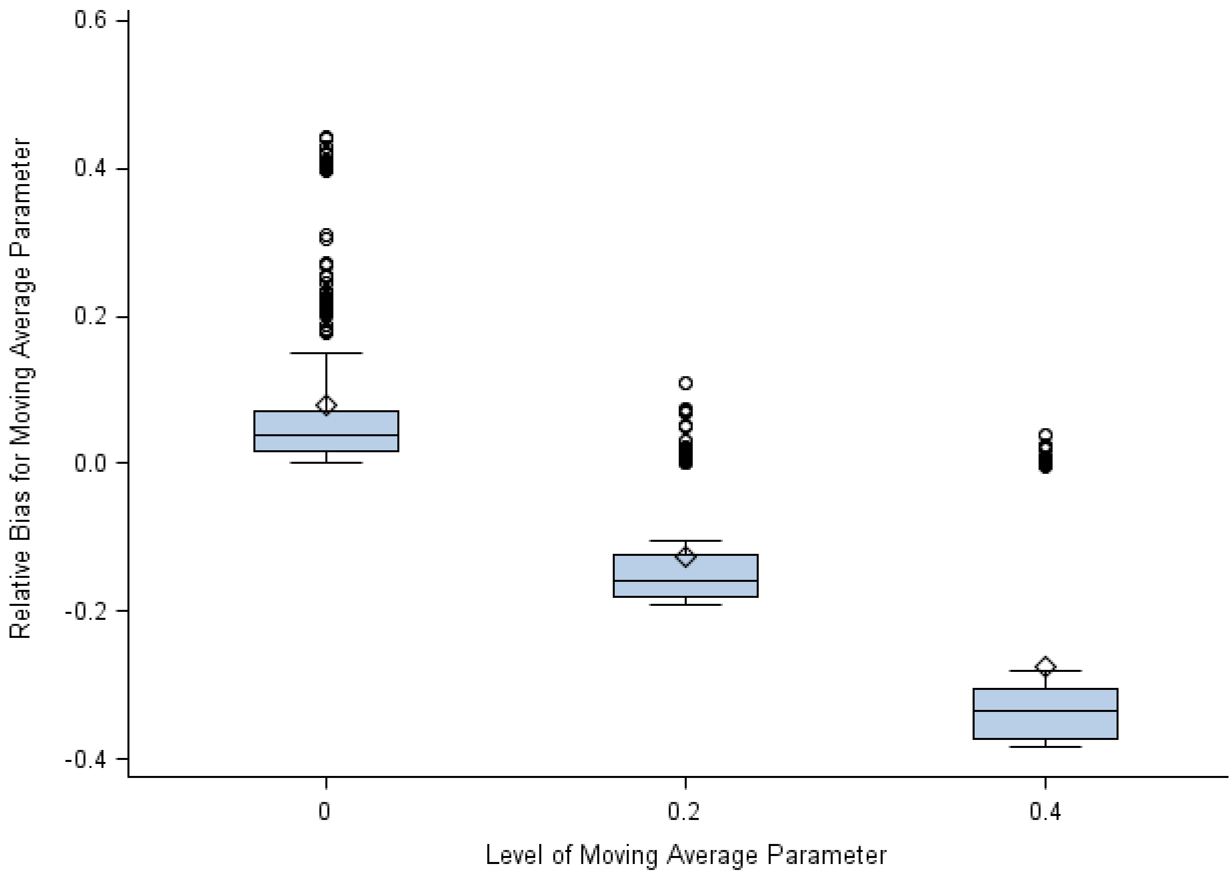

Figure 49. Box plot representing the relationship between the mean bias for the moving average parameter and the level of the moving average parameter.

GLM models were run to further explore the variability; the model, including five-way interactions explained 94% of the total variability. However, no considerable effects were found, therefore no further exploration was warranted for this data.

The data were trimmed to allow for further exploration of the RMSE values for the level-three variance for the phase effect. The distribution of the trimmed data is displayed in Figure 51. GLM models were run with the trimmed data and the results are further explained in Appendix A. Overall, the GLM models (including two-way interactions) explained 99% of the variability and revealed one medium or larger effect: the interaction of the number of primary studies with the variances of the error terms ($\eta^2 = 0.06$). This interaction was further investigated and the relationship was represented with line graphs. The line graphs depicted the interaction of the number of primary studies and the variance of the error terms with the trimmed RMSE values values for the level-three variance for the shift in level.

Further examination of the interaction with the RMSE values illustrated that for the less number of primary studies(10), the RMSE values tend to be greater than when the number of primary studies increased to 30, however the gap is even greater when the variances of the error terms is mostly at the upper levels.

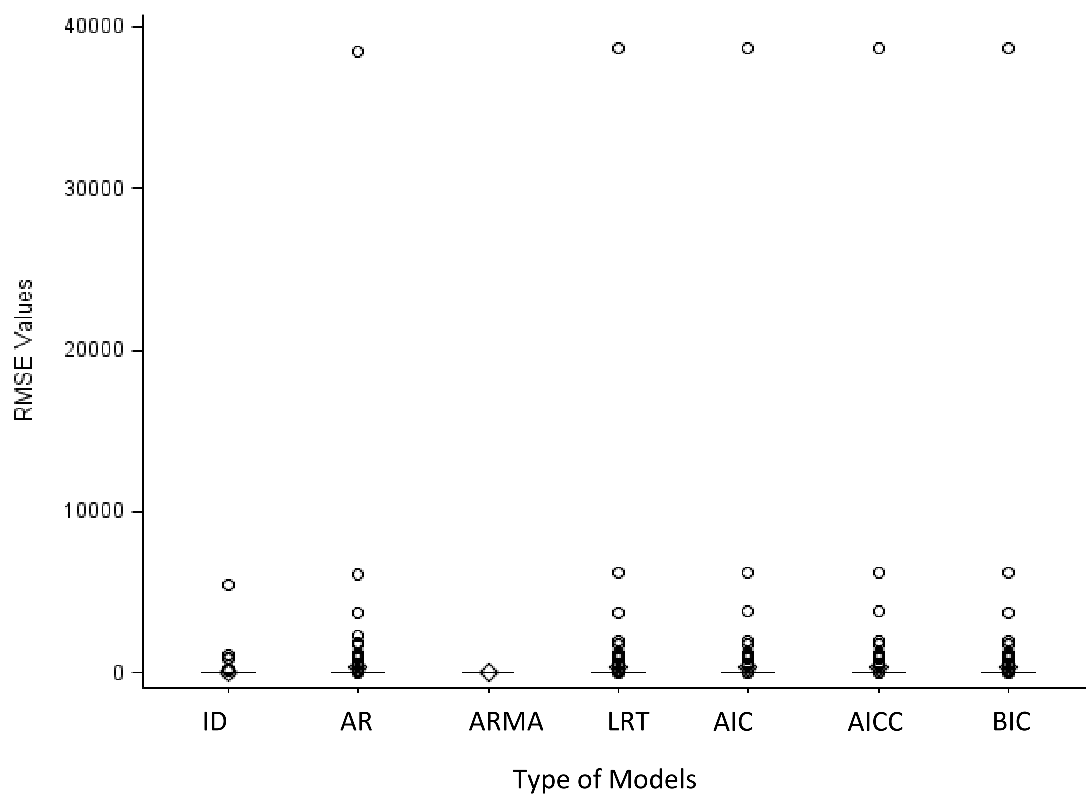

Figure 50. The distribution of the RMSE values for the phase effect for level-three variance across the seven models.

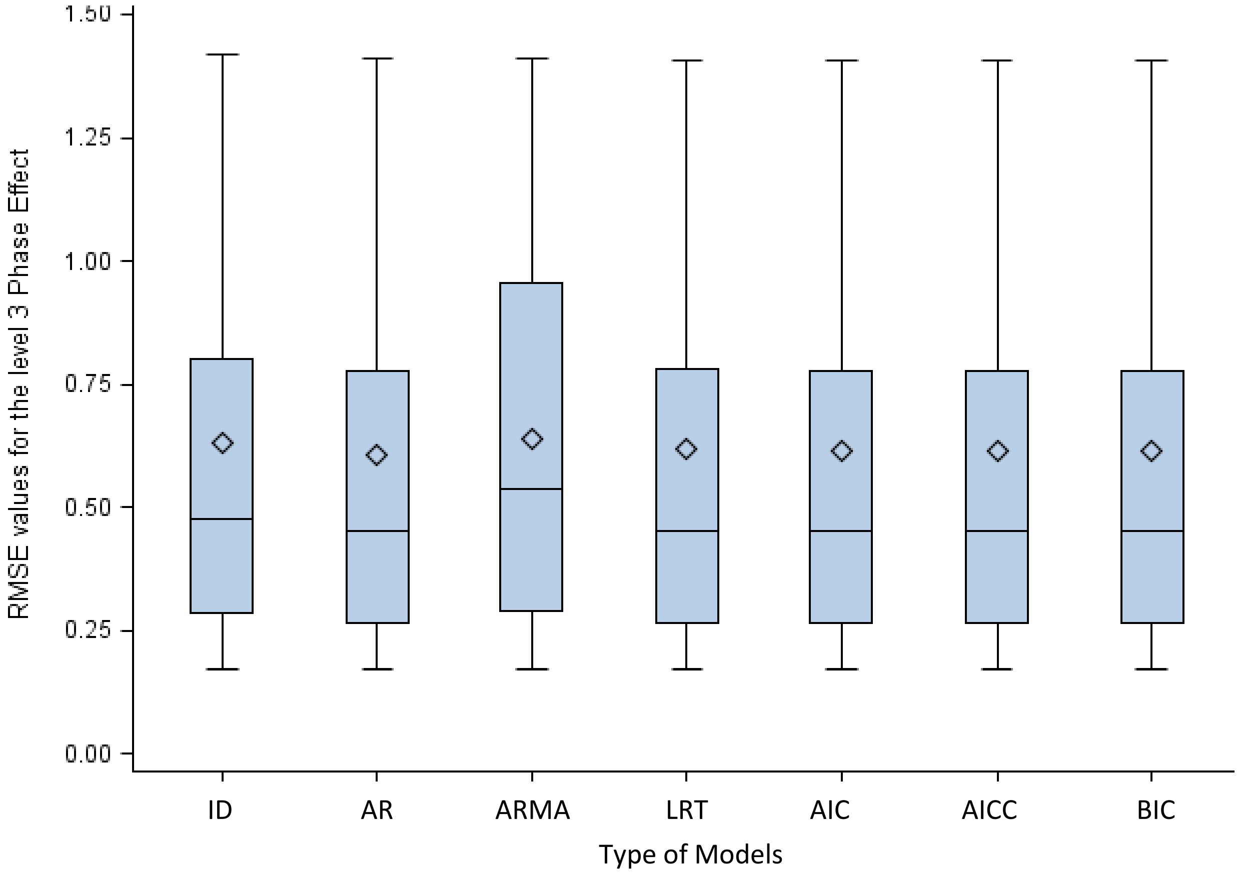

Figure 51. The distribution for the trimmed RMSE values for the level-three variance for the phase effect.

Level-three variance for the overall average treatment effect for the interaction effect (shift in slopes). The box plot illustrating the distribution for the RMSE values for the level-three variance for the interaction effect is displayed below in Figure 52. The means appear to be comparable across five of the models (the AR[1] model and each of the four fit index selected models). However, the mean RMSE value for the ID model was 2.06 (*SD* = 22.42) and the RMSE was smallest for the ARMA model (*M* = 0.07, *SD* = 0.04). To further explore the variability in the RMSE values for the level-three variance for the interaction effect, GLM models were run. The model, including five-way interactions, explained 93% of the total variability and none of the effects constituted a medium effect. Due to this, no further exploration was warranted for the original data.

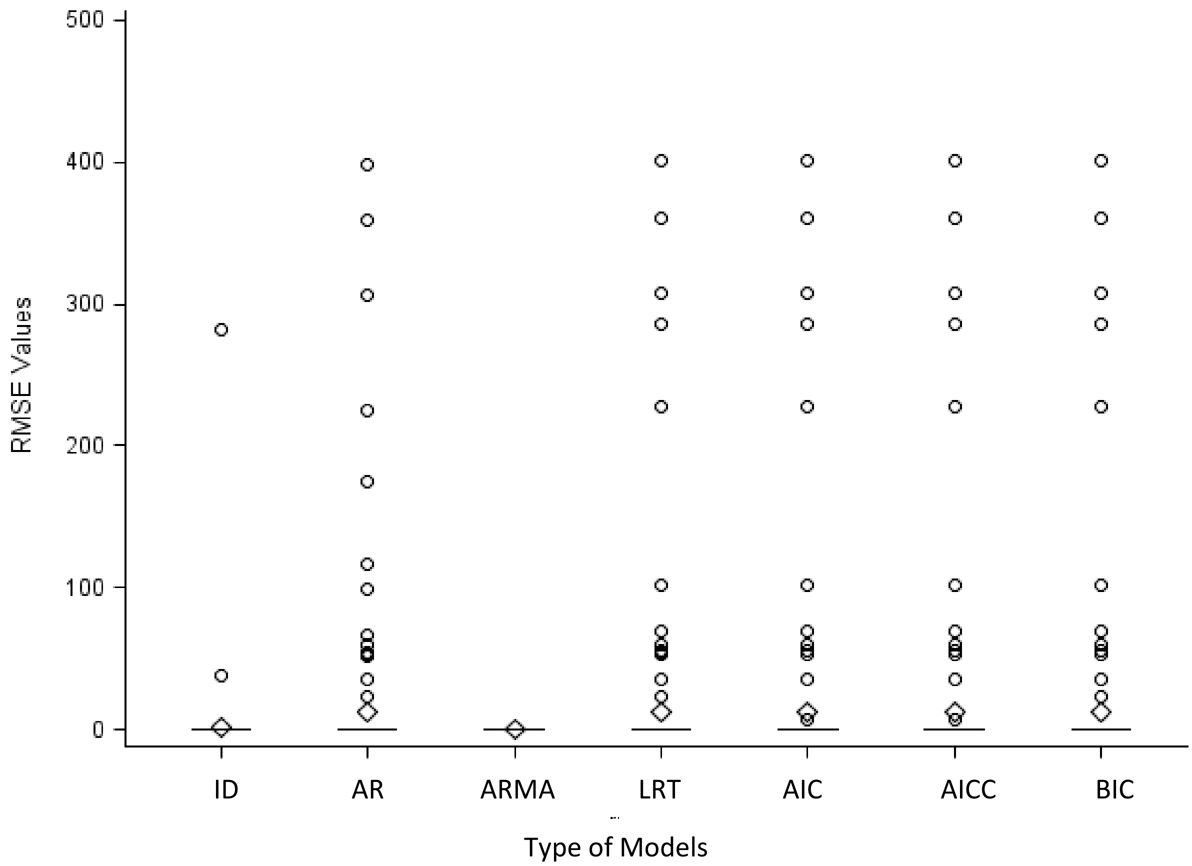

Figure 52. The distribution of the RMSE values for the level-three variance for the interaction effect (shift in slopes) across the seven models.

The RMSE values were then trimmed for further analysis of the RMSE values as a function of the seven models. The distribution for the trimmed RMSE values is displayed in Figure 53 below. GLM models were run to further investigate the relationship of the RMSE values with the study's design factors. The conclusions for the models were explained in Appendix A.

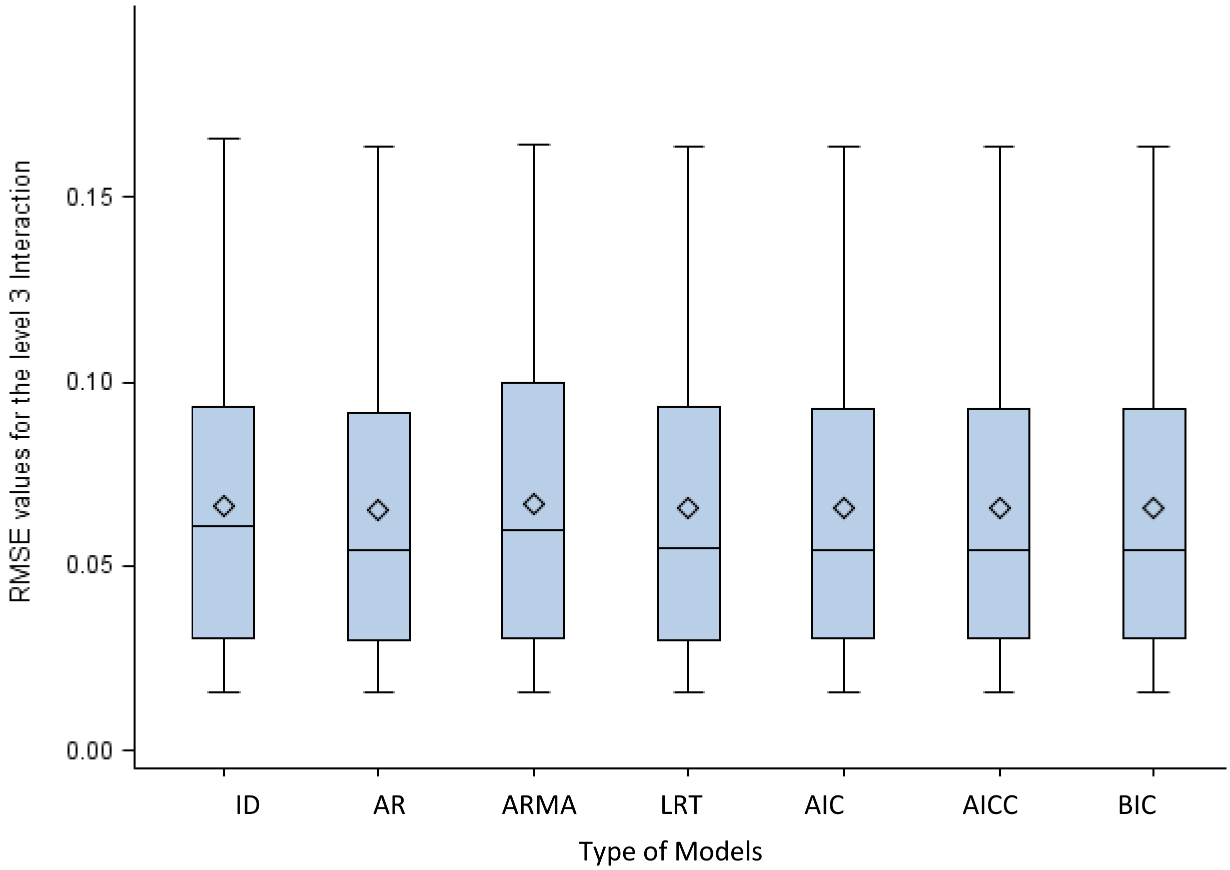

Figure 53. The distribution for the trimmed RMSE values for the level-three variance for the interaction effect.

Level-two variance for the average treatment effect for the phase (shift in level). The distribution for the level-two variance component for the phase effect is shown directly below in Figure 54. The means appear to be similar, with the type of model (η^2 = .011), indicating little variability between the types of model. The greatest mean for the RMSE was observed for the model selected by LRT (*M* = 295.32, *SD* = 1350.36), meanwhile the smallest mean RMSE value was noted for the AR(1) model (*M* = 0.40, *SD* = 0.21). To further examine the variability, GLM models, including 5-way interactions were run. No significant or medium effects were found, no further analyses are warranted.

The RMSE values were then trimmed for further analysis of the RMSE values as a function of the seven models. The distribution for the trimmed RMSE values is displayed in Figure 55. GLM models were run to further investigate the relationship of the RMSE values with the study's design factors. The findings for the models are explained with greater detail in Appendix A. In summary, the model, including third order interactions, explained 95.9% of the variability. There were three medium or larger effects: number of participants (η^2= 0.12), number of studies to be included in the meta-analysis (η^2= 0.22), the variance of the error terms (η^2= 0.43).

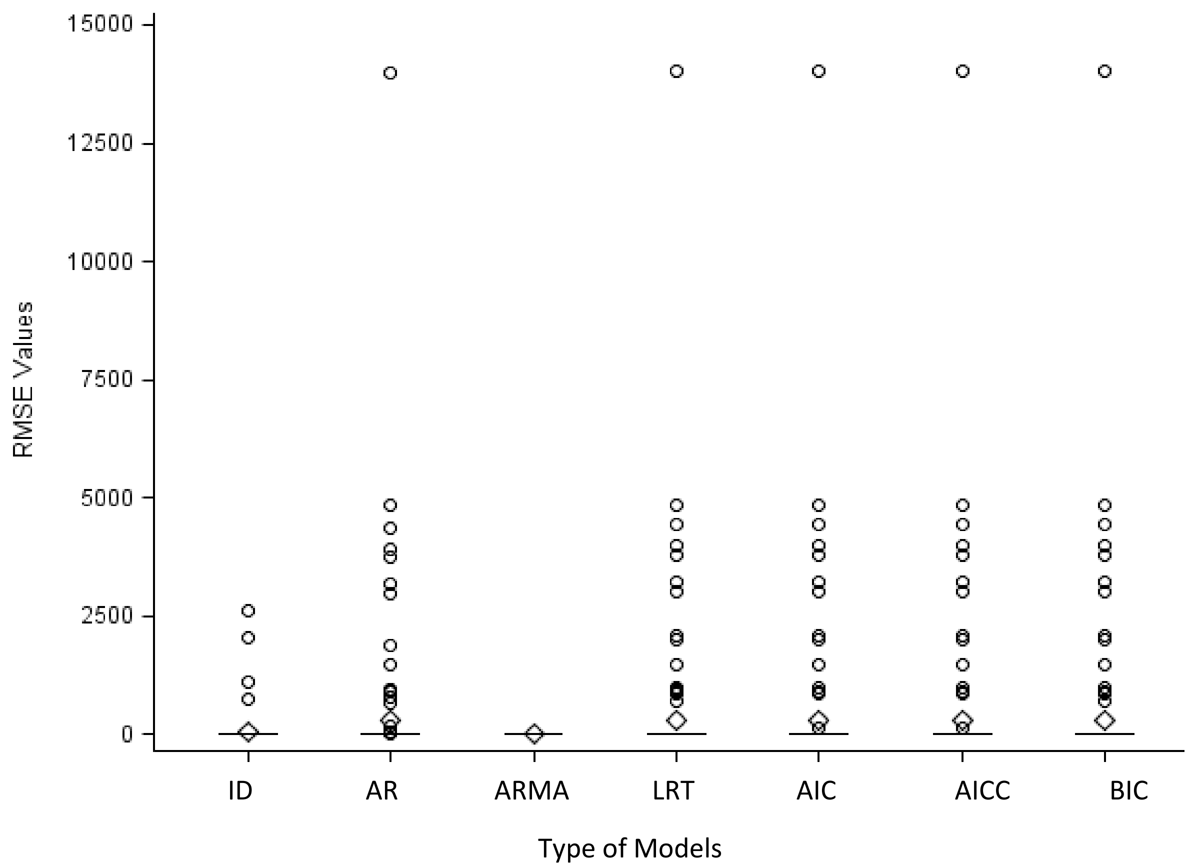

Figure 54. Box plot illustrating the distribution of the RMSE values for the level-two variance components for the phase effect across the seven models.

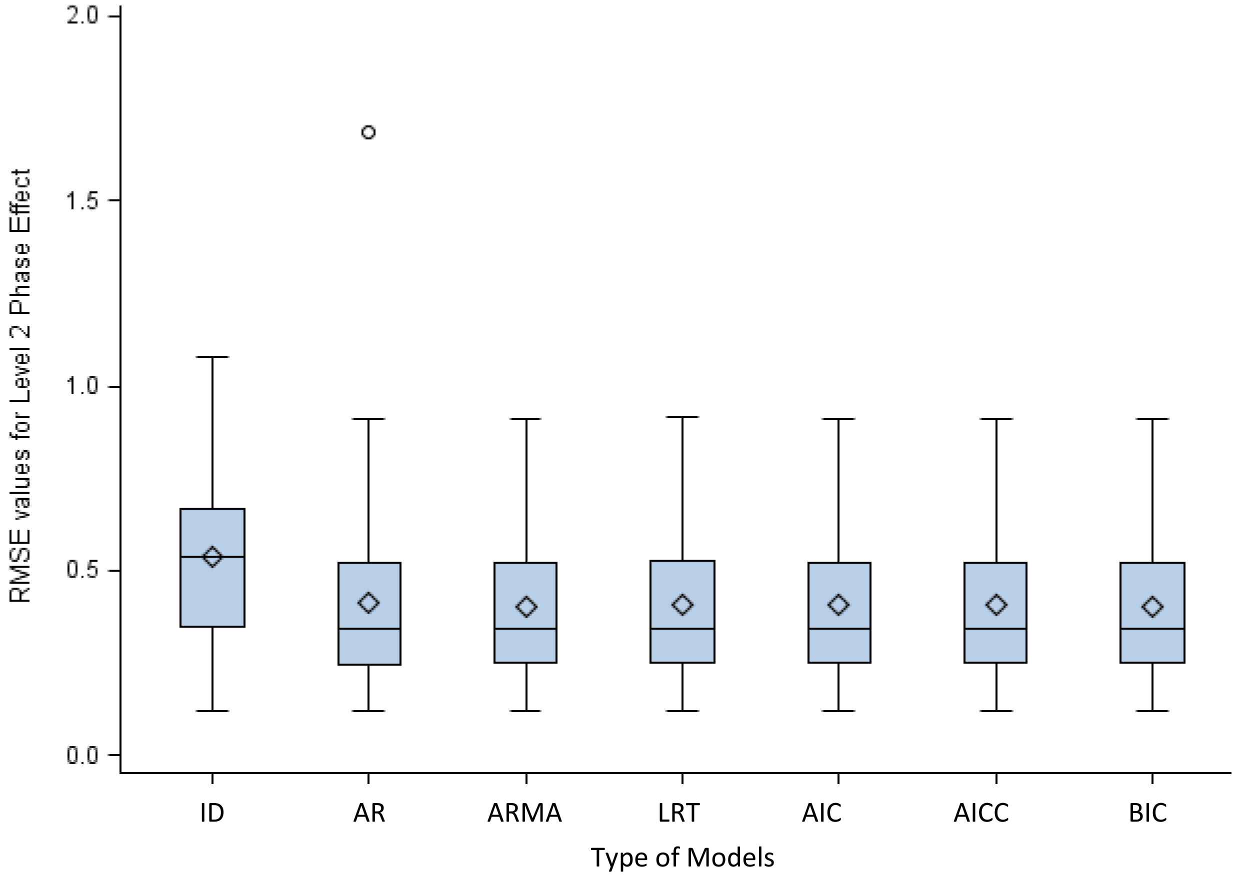

Figure 55. The box plot illustrating the distribution for the trimmed RMSE values for the level-two phase effect.

Level-two variance for the overall average treatment effect for the interaction (shift in slopes). The distribution for the level-two variance components for the shifts in slopes is displayed in the box plot below (see Figure 56). The largest mean was observed for the model selected by the LRT (M = 12.77, SD = 54.84), conversely, the smallest mean was for the AR model (M = 0.07, SD = 0.04). The type of model (η^2 = 0.017), indicating a menial amount of variability between the type of models. GLM models were run to further analyze the variability. The model, including five-way interactions explained 97% of the total variability. None of the effects met the criteria for a medium, therefore no further exploration was warranted.

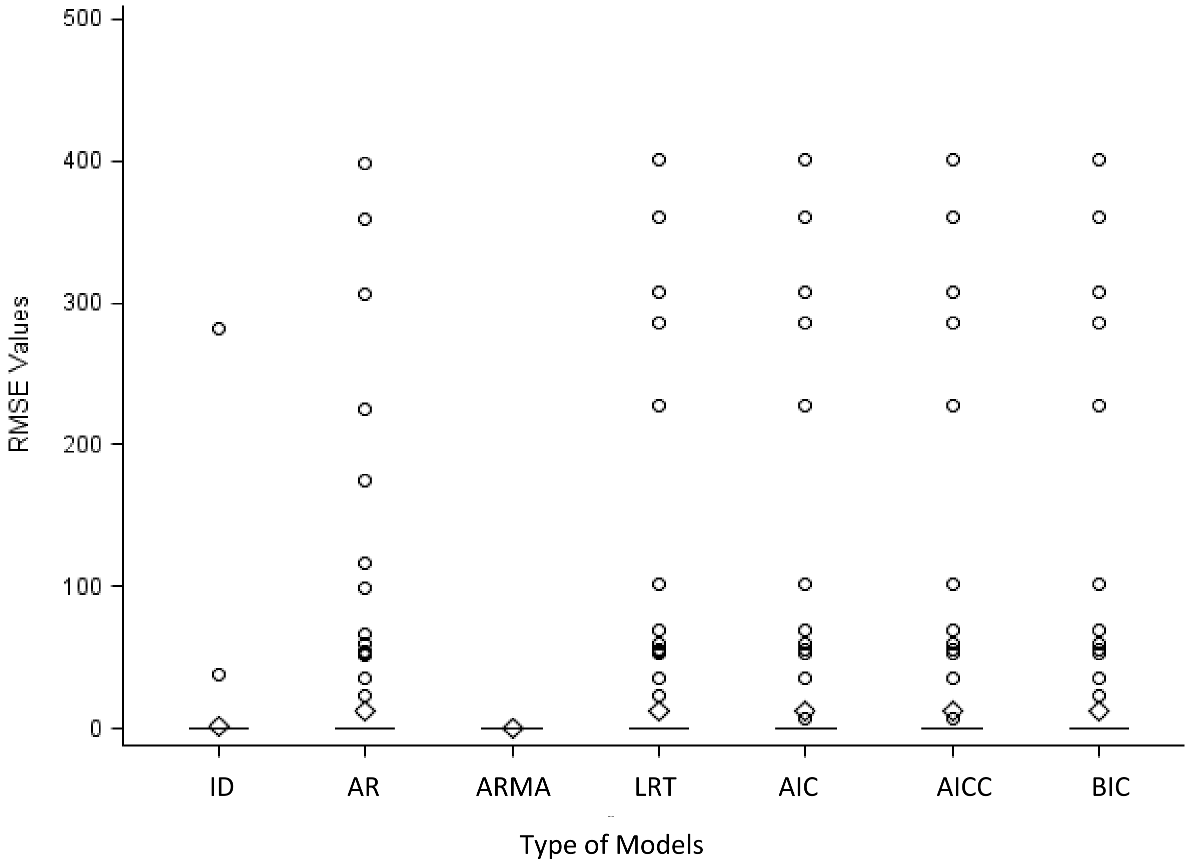

Figure 56. Box plot illustrating the distribution of the RMSE values for the level-two variance components for the interaction effect across the seven models.

The data were then trimmed to further investigate the variability in level-two variance for the interaction effect; the graph for the trimmed distribution is displayed below in Figure 57. The graph shows that the means are comparable across most of the models, with the ID model having a slightly larger mean. The greatest variability was also observed for the ID models, and againk similar variability was noticed for the remainder of the models. GLM models were then run to determine if there were any medium or larger effect for each of the design factors and combinations of these factors. The results are presented and explained in great detail in the Appendix A.

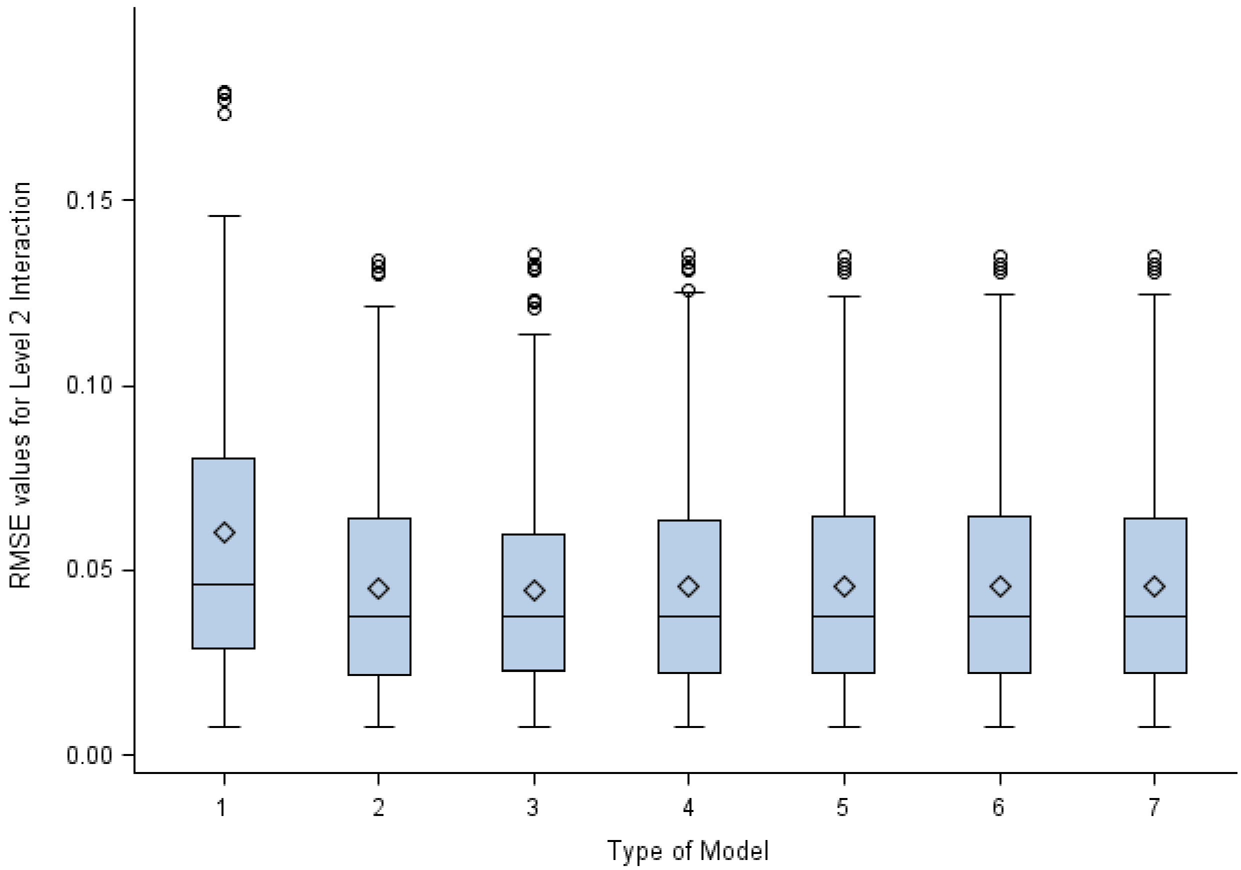

Figure 57. The distribution for the trimmed RMSE values for the level-two variance for the interaction effect.

Level-one or Residual Variance. Figure 58 below displays the distribution of the RMSE values for the level-one variance across the seven models. The largest mean RMSE value was observed for the model selected by the AR model, *M* = 11.97, *SD* = 129.02, conversely, the smallest mean was for the ID model (*M* = 0.10, *SD* = 0.05). The type of model (η^2 = 0.017), indicating a menial amount of variability between the type of models. GLM models were run to further analyze the variability. The model, including five-way interactions explained 97% of the total variability. None of the effects met the criteria for at least a medium effect; therefore no further exploration was warranted on the original data.

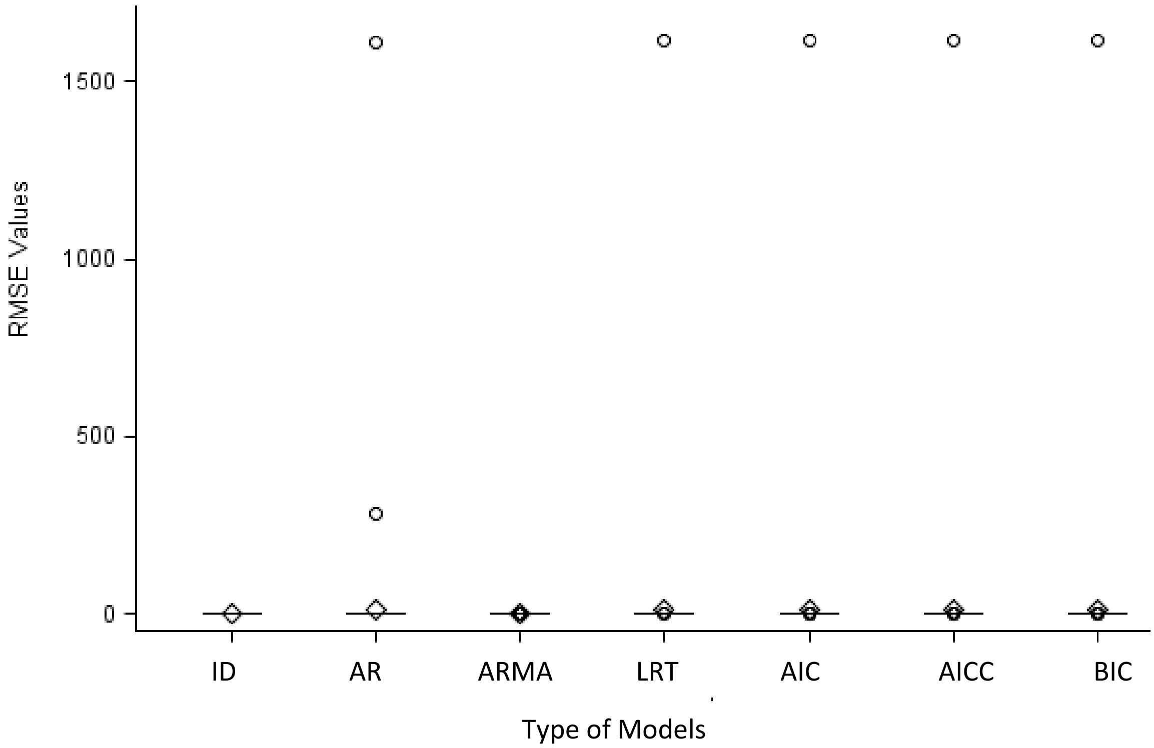

Figure 58. Box plot showing the distribution of the RMSE values for the level-one variance across the seven models.

Due to the extreme variability observed in several points for the RMSE values for the level-one variance, the data were trimmed and six data points were removed. The resulting distribution of the trimmed RMSE values for the level-one variance is displayed below in Figure 59. The figure displays that the mean RMSE values varied across the different types of models. The largest mean RMSE values (M = 0.36, SD = 0.29) for the level-one variance was observed for ARMA (1,1) model, while the smallest mean RMSE values (M = 0.10, SD = 0.06) was noted for the ID model. To further explore the variability of the trimmed RMSE values for the level-one variance, GLM models were used. The findings are further explicated in the Appendix A.

Overall, The resulting model included 3-way interactions and explained 98% of the total variability. The following medium effects were found: the series length or number of observations ($\eta^2 = 0.06$), the interaction of the level of the autocorrelation parameter and the type of model ($\eta^2 = 0.07$), the interaction of the variances of the error terms and the type of model ($\eta^2 = 0.11$).

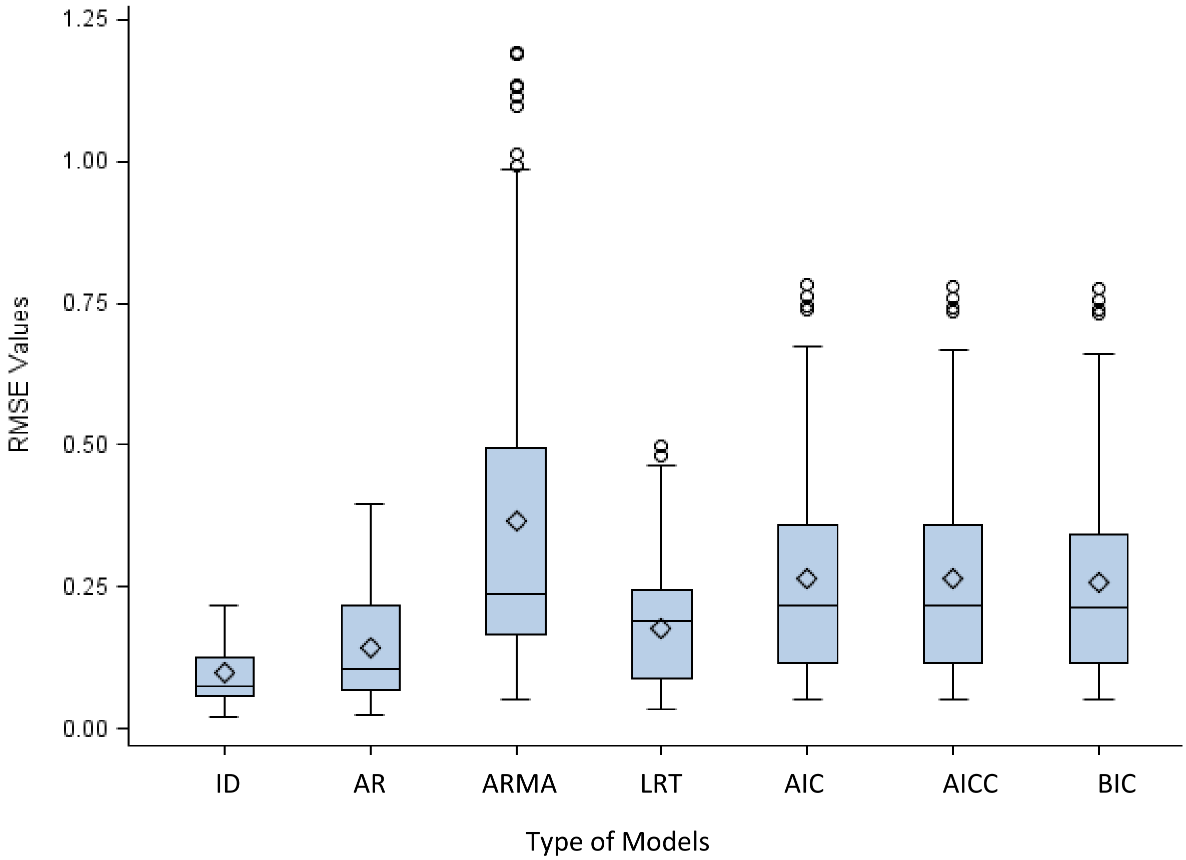

Figure 59. Box plot showing the distribution of the trimmed RMSE values for the level-one variance across the seven models.

Autocorrelation Parameter. The box plot below (see Figure 60 below) illustrates the distribution of the RMSE values for the autocorrelation parameter across the six models (the autocorrelation parameter was estimated to be 0 for the ID model). The plots indicated that the smallest mean for the RMSE values for the autocorrelation parameter is for the first-order autoregressive model (M = 0.05, SD = 0.02), while the largest mean for the RMSE values was observed for the moving average model (M = 0.33, SD = 0.24). To further explore the variability in the RMSE values for the autocorrelation parameter, GLM models were run. The models, which included two-way interactions, explained 98% of the total variability. The model revealed that there was one effect that met the aforementioned criteria for a medium effect: the interaction of type of model and the level of the autocorrelation parameter ($\eta^2 = 0.28$).

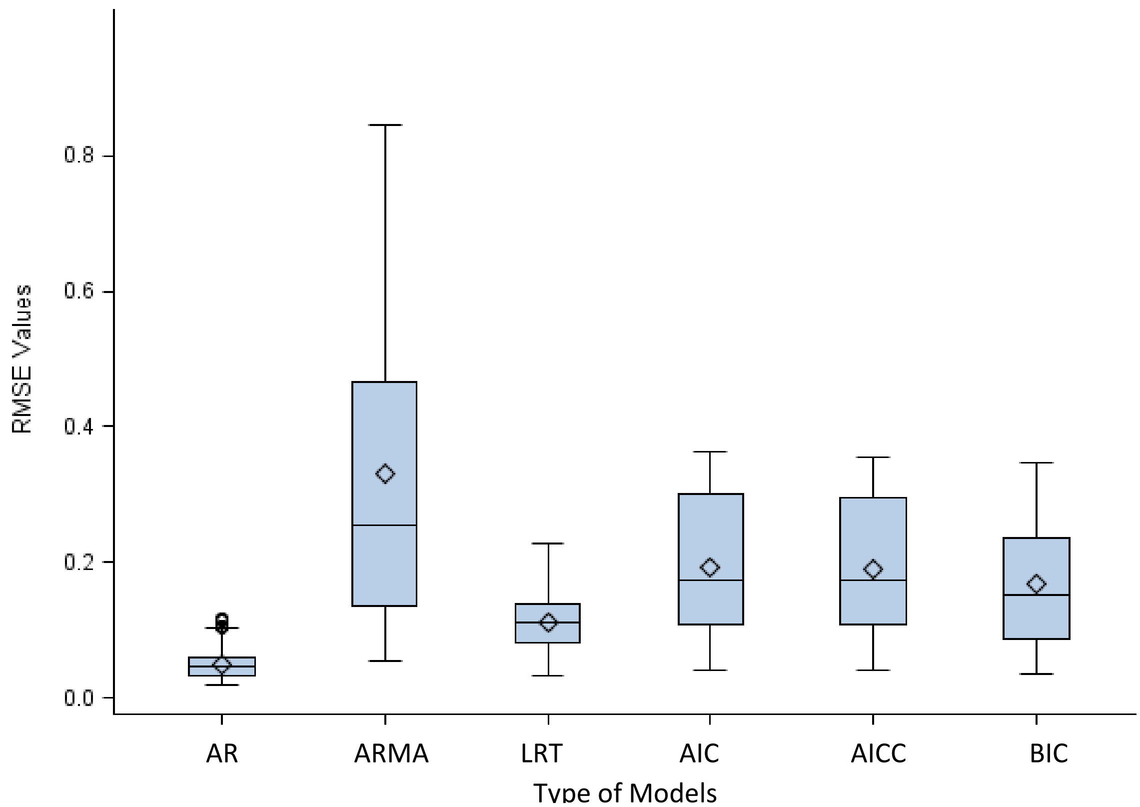

Figure 60. Box plot showing the distribution for the RMSE values for the autocorrelation parameter across the six models.

To further examine the significant effect, line graphs (see Figure 61 below) were created for the relationship of the RMSE values for the autocorrelation parameter and the interaction effect between the level of the autocorrelation parameter and type of model. The line graph indicated that RMSE values were similar for the models selected by three of the fit indices (AIC, AICC, and BIC) across the various levels of the autocorrelation parameter. For the models selected by the LRT, the RMSE values were slightly smaller, and smallest for the AR(1) model across the various levels of the autocorrelation parameter. The largest RMSE values were observed for the ARMA(1,1) model with a mean of 0.15 (*SD* = 0.08) for ρ = 0.4, mean of 0.30 (*SD* = 0.13) for ρ = 0.2, and finally, mean of 0.75 (*SD* = 0.07) for ρ = 0.0.

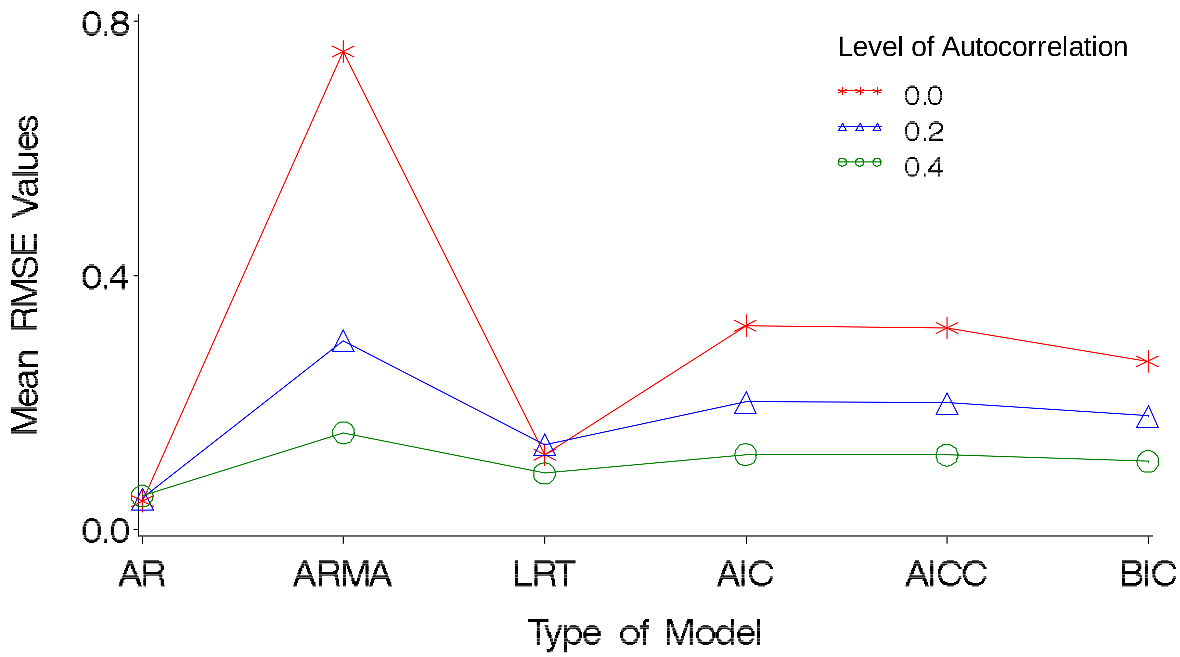

Figure 61. Line graph depicting the relationship between the mean RMSE values for the autocorrelation parameter and the interaction effect of type of model and the level of the autocorrelation parameter.

First-Order Autoregressive Moving Average Parameter. The box plot displayed below in Figure 62 depicts the distribution of the RMSE values for the moving average parameter across the five models (the moving average parameter was estimated as zero for the ID and the AR models). The smallest mean for the RMSE values were observed for the LRT models (*M* = 0.16, *SD* = 0.13) and the largest mean for the RMSE values was for the ARMA(1,1) model (*M* = 0.21, *SD* = 0.14). GLM models were run to further explore the variability of the RMSE values, and their relationship with the combination of design factors.

The model, including two-way interactions explained 97% of the total variability, and revealed that there were two significant (medium or larger) effects: the level of the autocorrelation parameter ($\eta^2 = 0.26$) and the interaction between the level of the moving average parameter and the type of model ($\eta^2 = 0.45$). Line graphs were then created to further examine the relationship of these effects with the RMSE values for the moving average parameter.

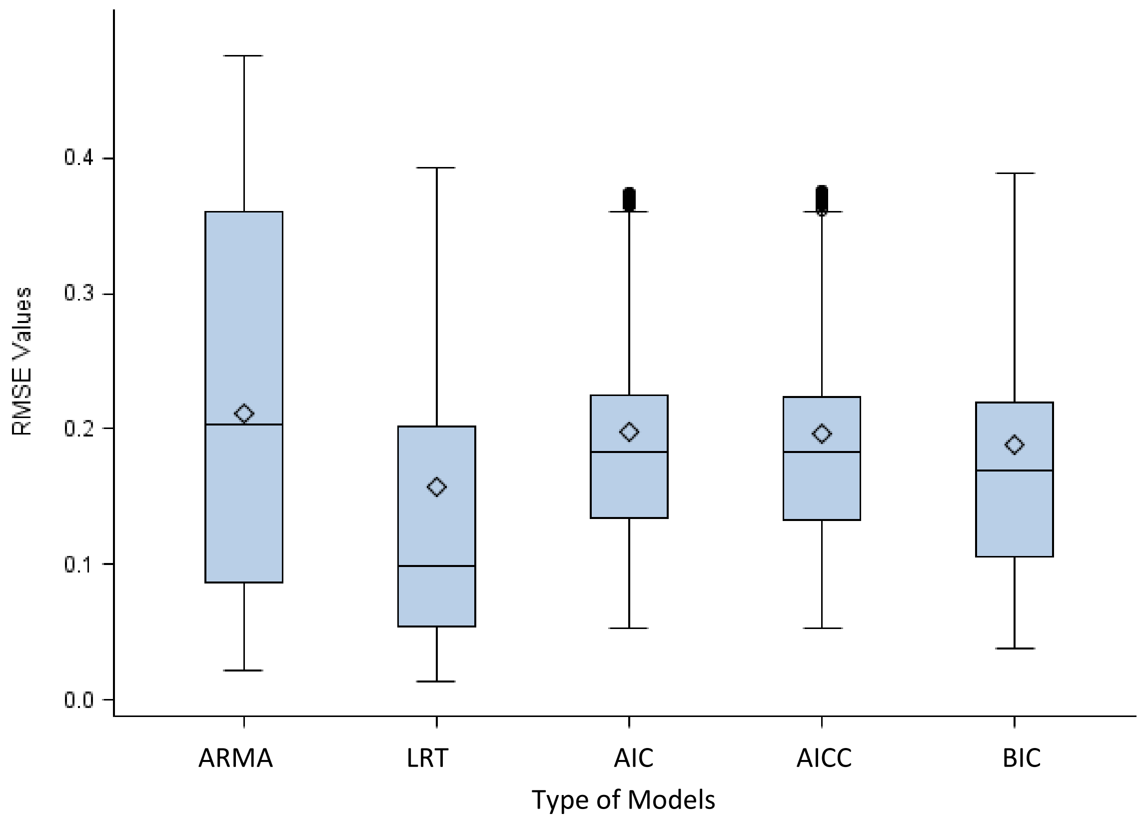

Figure 62. Box plot illustrating the distribution of the RMSE values for the moving average parameter across the five models.

Line graphs were then created to explore the association of the mean for the RMSE values and these effects. Figure 63 below depicts the relationship between the mean RMSE values and the interaction between the level of the moving average parameter and the type of model. The graph shows that for the models selected by the various fit indices that the mean

RMSE values increase as the level of the moving average parameter increase. However, the inverse relationship was observed for the ARMA(1,1) model, the mean RMSE values increased as the level of the moving average decreased. Specifically, the mean RMSE value was smallest for θ = 0.4 (*M* = 0.07, *SD* = 0.04) and largest for θ = 0.0 (*M* = 0.30, *SD* = 0.11).

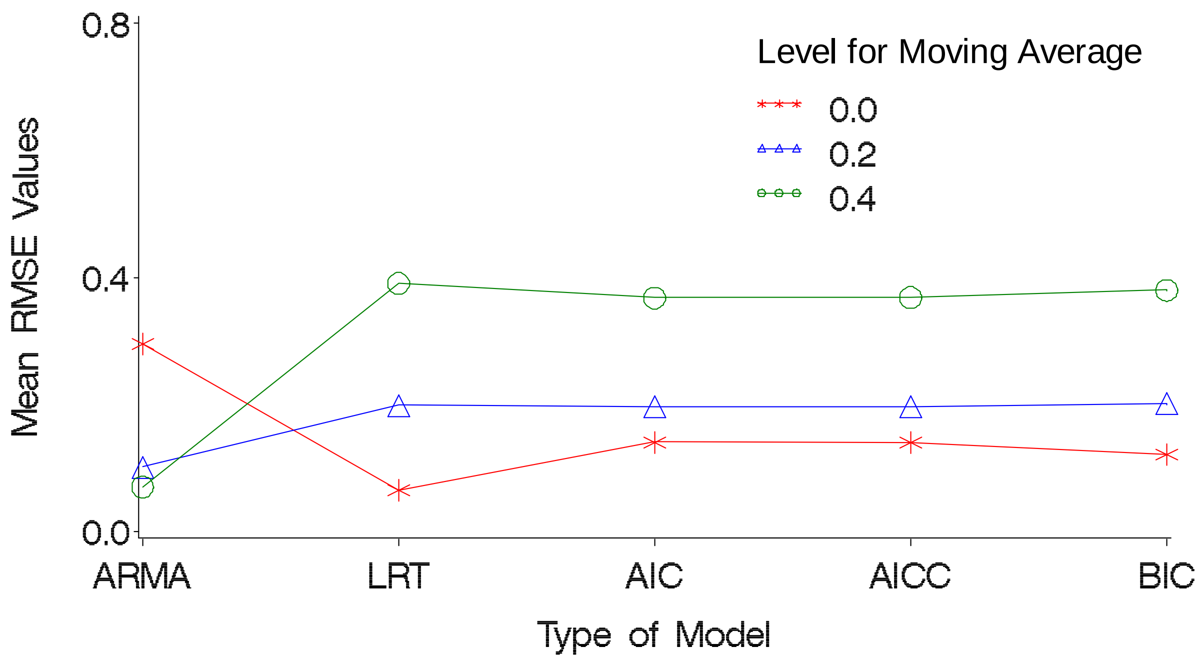

Figure 63. Line graphs for the association of the mean RMSE values for the moving average parameter and the interaction between level of the moving average parameter and the type of model.

Confidence Interval Coverage

The confidence interval coverage will be explored in the next section for all of the variance components. These components included the level-three variance for both the phase effect (shift in level) and the interaction effect (shift in slopes), the level-two variance for both the both the phase effect (shift in level) and the interaction effect (shift in slopes), the level-one or residual variance, the autocorrelation parameter, and finally, the moving average parameter.

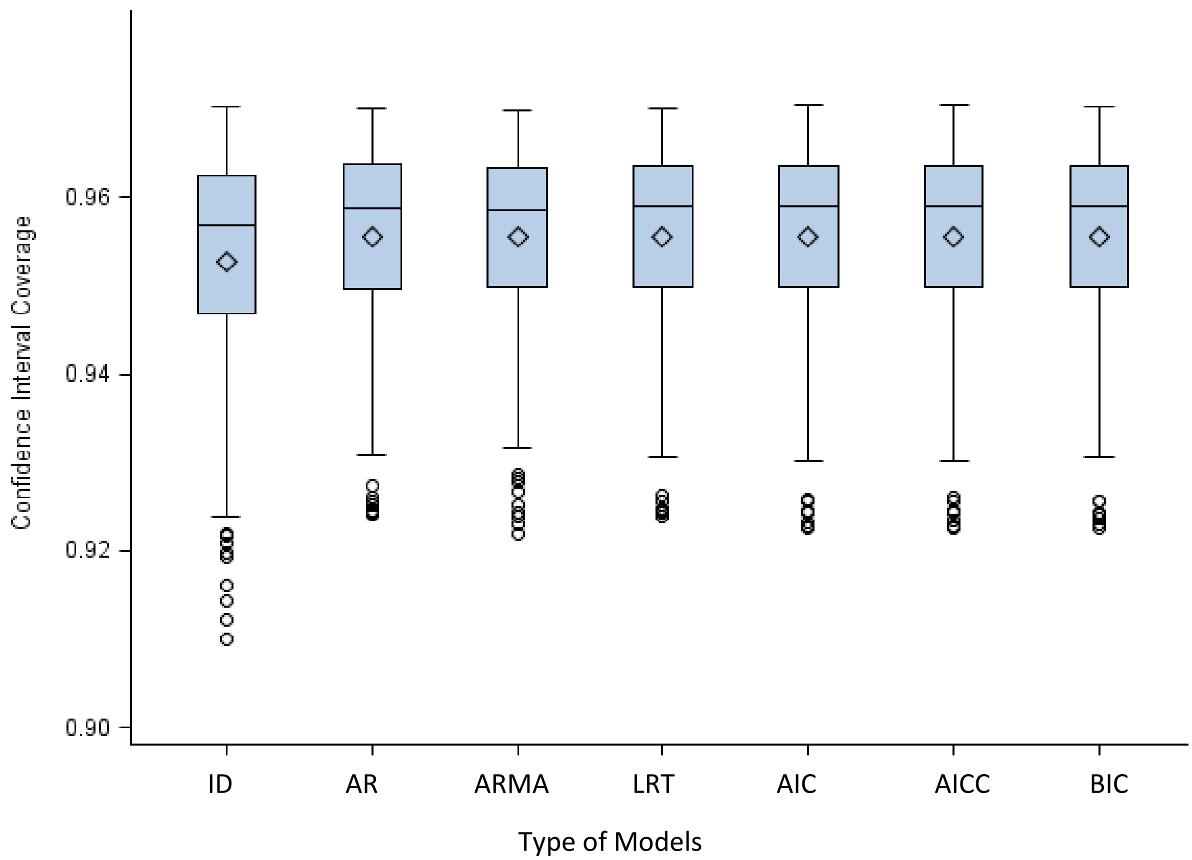

Figure 64. Box plots for the distribution of the coverage for the level-three variance for the phase effect (shift in levels) across the seven models.

Level-three variance for the overall average treatment effect for the phase (shift in level). Figure 64 above shows the distribution for the confidence interval coverage for the level-three variance for the phase effect across the seven models. The means for the confidence interval coverage appear to be comparable ($\eta^2 = 0.007$), further indicating the lack of variability across the seven models. The largest mean was for the model selected by LRT (*M* = 0.9556, *SD* = 0.011) and the ID model had the smallest mean interval coverage (*M* = 0.9528, *SD* = 0.014).

To further examine the variability in the confidence interval coverage, GLM models were run. The model, including two-way interactions, explained 95% of the total variability, and indicated that two combinations of design factors were medium effects: the number of primary studies included in the meta-analysis ($\eta^2 = .07$) and the interaction effect between the number of participants and the variances of the error terms ($\eta^2 = .09$).

Box plots (see Figure 65 and 66) were then created to explore the relationship with each of these medium effects with the outcome of interest and the confidence interval coverage for the level-three variance for the shift in level. First, in Figure 65, the graph illustrates the relationship between the mean interval coverage and the number of primary studies to be included in the meta-analysis, revealing that as the number of primary studies to be included in the meta-analysis increased from 10 (*M* = .95, *SD* = 0.013) to 30 (*M* = 0.96, *SD* = 0.008), the mean confidence interval coverage also increased.

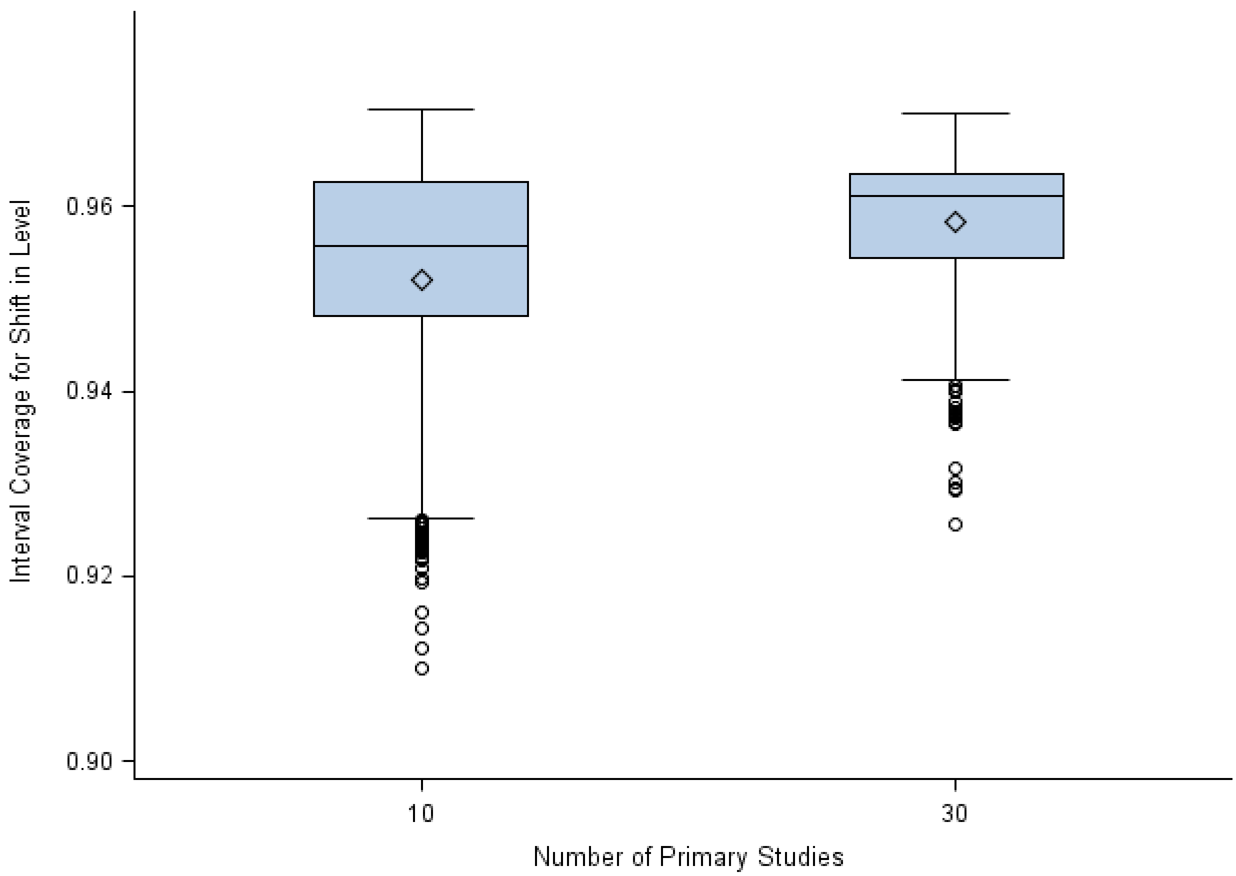

Figure 65. Box plot depicting the relationship between the mean confidence interval coverage for level-three variance for the phase effect and the number of primary studies to be included in the meta-analysis.

The association of the interval coverage and the interaction of the variances of the error terms and the number of participants in each study was then illustrated in Figure 66 below. The graph indicates that the effect of the number of participants (from 4 to 8) on the mean interval coverage is dependent upon the variances of the error terms. Specifically, when most variances of the error terms are at the upper levels for the shift in level then there is only a slight increase in mean as the number of participants shift from 4 (M = 0.96, SD = 0.004) to 8(M = 0.96, SD = 0.003). However, there is a greater increase in the mean interval coverage when the most of the variances of the errors are at level-one for the phase effect as the number of participants shift from 4 (M = 0.94, SD = 0.010) to 8 (M = 0.96, SD = 0.006).

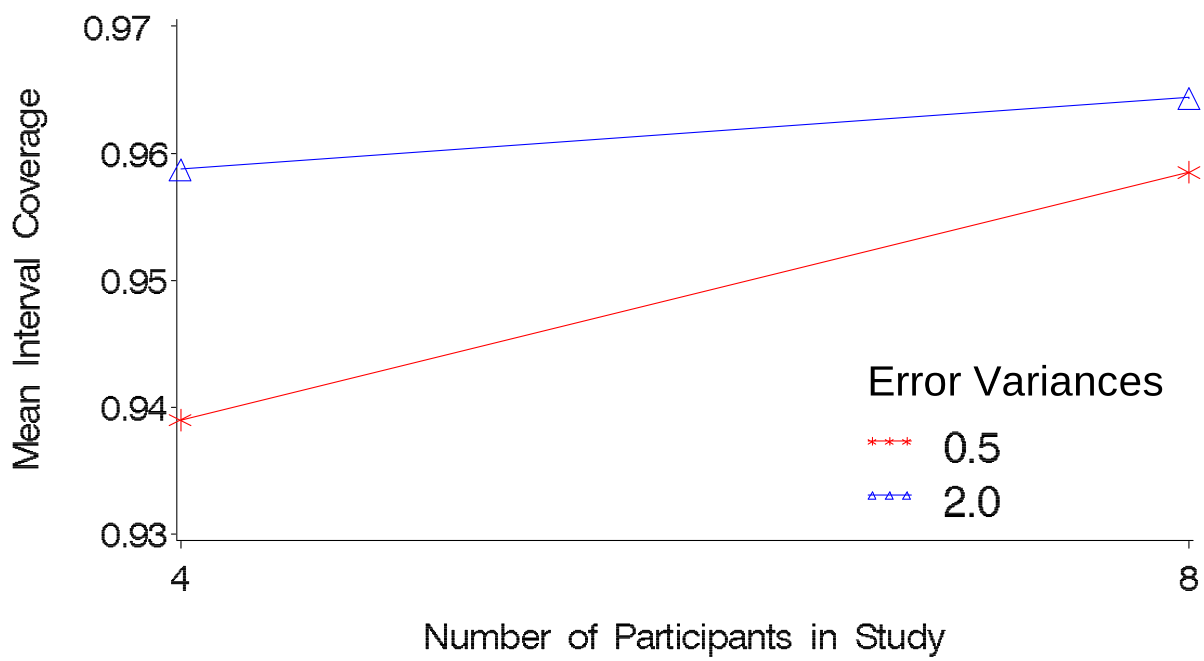

Figure 66. Line graph illustrating the relationship between the mean confidence interval coverage for the level-three variance for the phase effect and the interaction between the variances of the error terms and the number of participants in study.

Level-three variance for the overall average treatment effect for the interaction effect (shift in slopes). The distribution for the confidence interval coverage for the level-three variance for the interaction effect is shown in Figure 67 below. The figure shows that the variability across the models are minimum ($\eta^2 = .014$). The largest mean interval coverage was for the model selected by the BIC fit index (*M* = 0.952, *SD* = 0.020); the smallest mean interval coverage was for the ID model (*M* = 0.944, *SD* = 0.028). To further explore the variability in the mean interval coverage, GLM models were run.

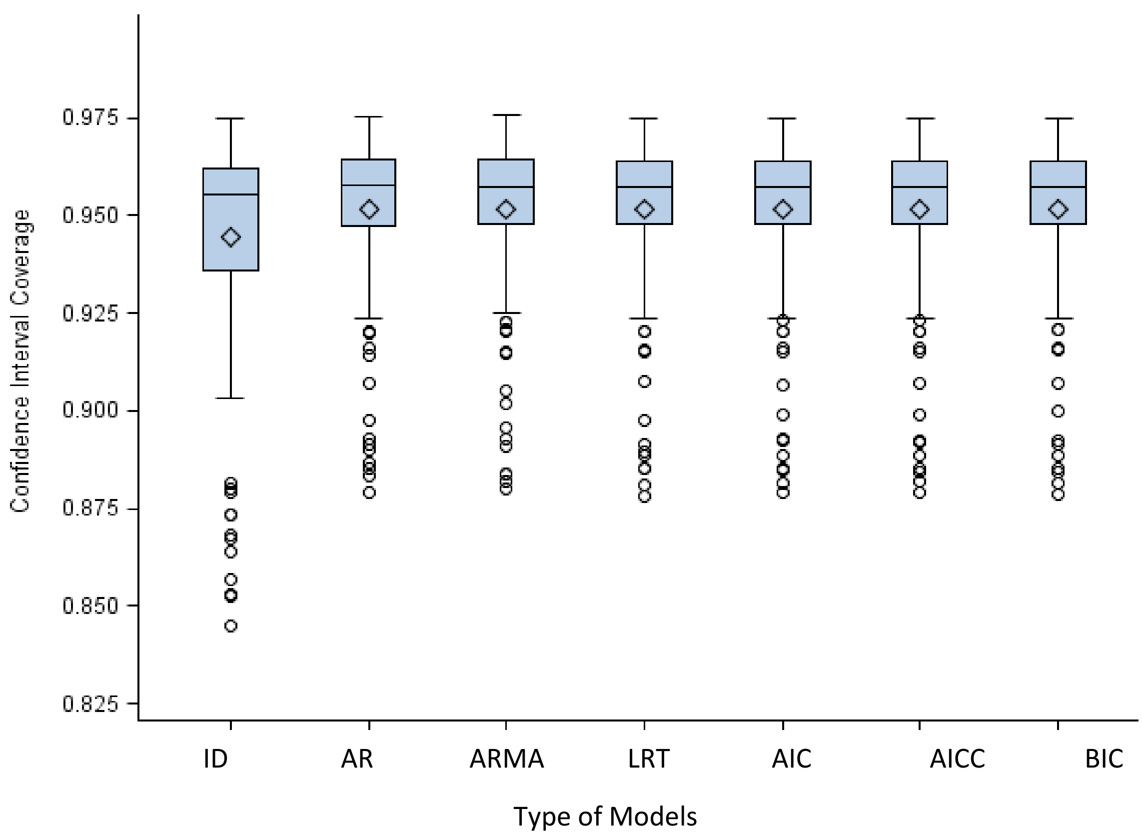

Figure 67. Box plot showing the distribution of the confidence interval coverage for the level-three variance for the interaction effect (shift in slopes) across the seven models.

The results of the model including two-way interactions, explained 95% of the total variability, reported two effects that met the aforementioned criteria for being described as a medium effect. One of the effects was the interaction effect between the series length and the variances of the error terms ($\eta^2 = .11$). The other medium effect was the number of participants in the study ($\eta^2 = .14$). Line graphs were then constructed to further illustrate the relationship between the mean interval coverage and these effects. The graph below (see Figure 68) represents the relationship between the mean interval coverage and the interaction between the series length and the variances of the error terms. Furthermore, the lines illustrate that the effect or the impact of the series length on the mean confidence interval coverage depends on the variances of the error terms. Specifically, the increase in the mean confidence interval coverage is greater when the series length is increases from 10 (*M* = 0.92, *SD* = 0.025) to 20 (*M* = 0.96, *SD* = 0.007) when most of the variance is at level-one. Conversely, when most of the variance is at the upper levels, the effect of the series length increasing from 10 (*M* = 0.954, *SD* = 0.010) to 20 (*M* = 0.963, *SD* = 0.005) on the mean interval coverage is minimal.

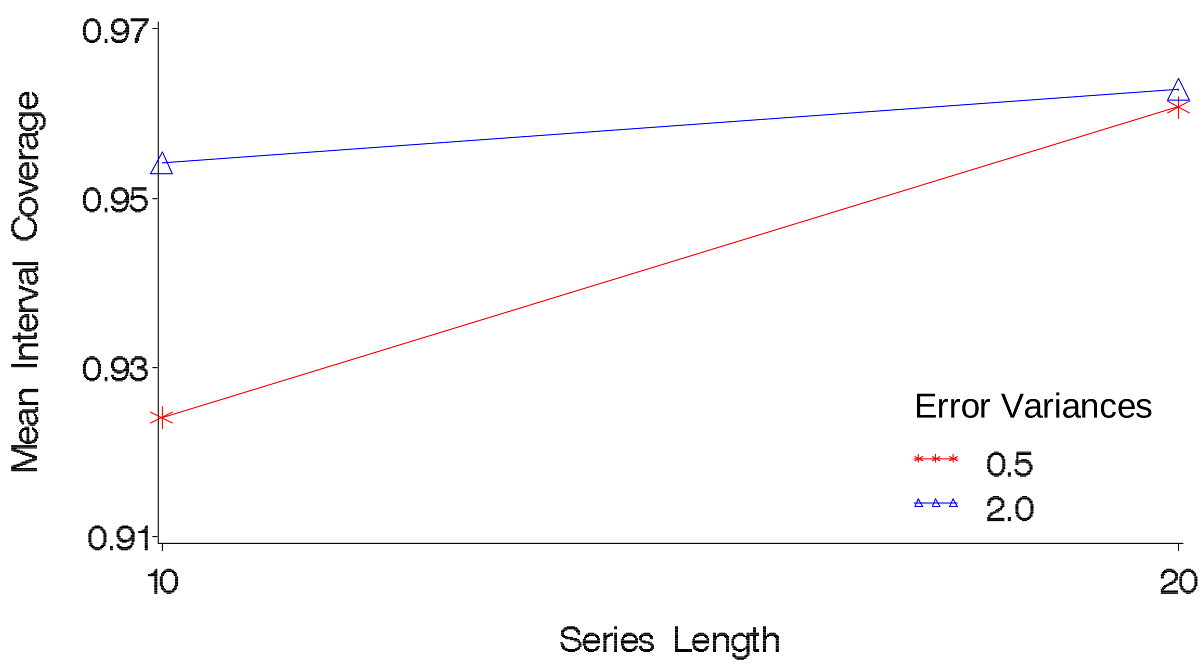

Figure 68. Line graphs depicting the relationship between the mean interval coverage for the level-three variance of the interaction effect (shift in slopes) and the interaction of the variances of the error terms and the series length.

Figure 69 illustrates the relationship of the outcome of interest with the second medium effect, specifically, the relationship between the mean confidence interval coverage for the level-three variance for the interaction effect and the number of participants included in a primary study. The graph reveals that as the number of participants increased from 4 (M = 0.94, SD = 0.025) to 8 (M = 0.96, SD = 0.011), then the mean confidence interval width also increases.

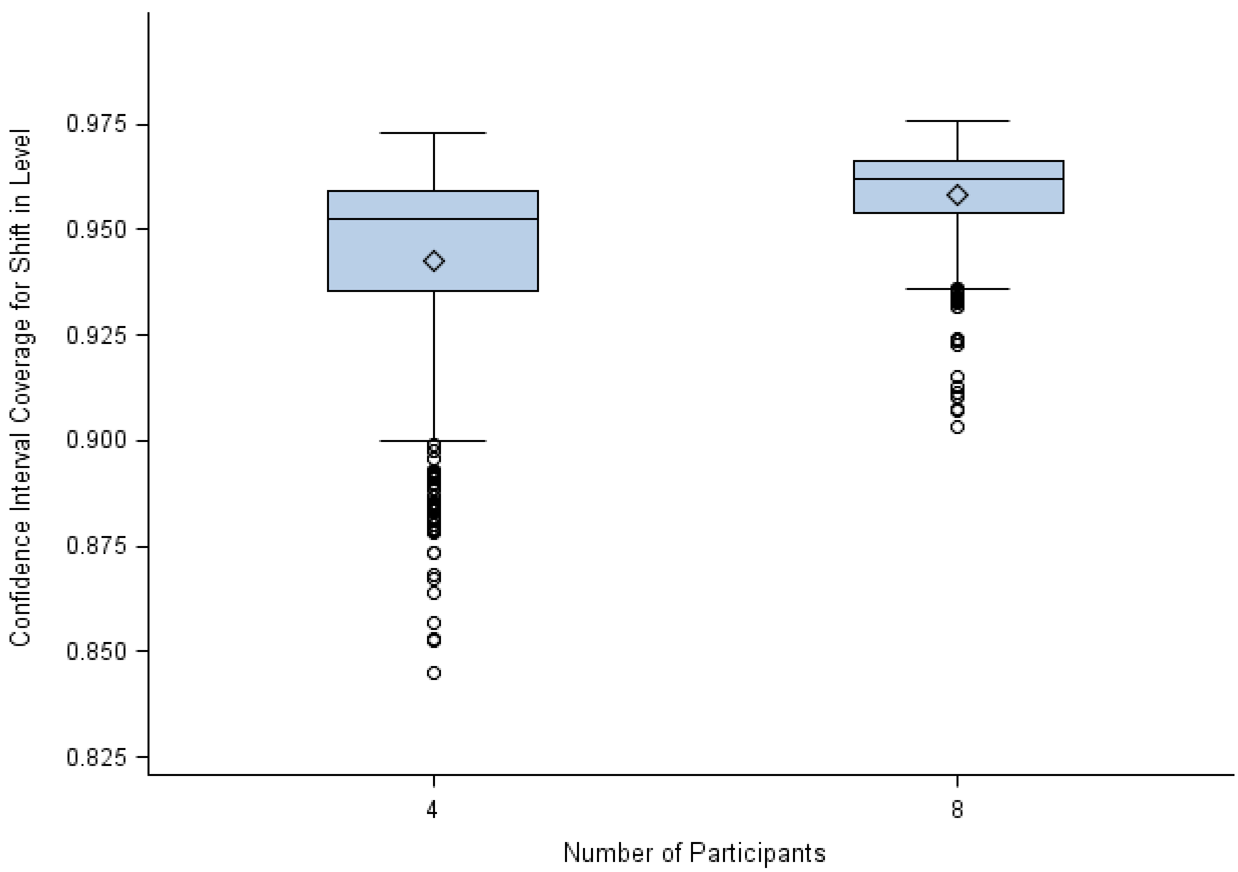

Figure 69. Box plot depicting the association between the mean interval coverage for the level-three variance for the interaction effect (shift in slopes) and the number of participants in a particular study.

Level-two variance for the average treatment effect for the phase (shift in level). The distribution of the confidence interval coverage for the level-two variance for the phase effect (shift in level) in shown in Figure 70 below.

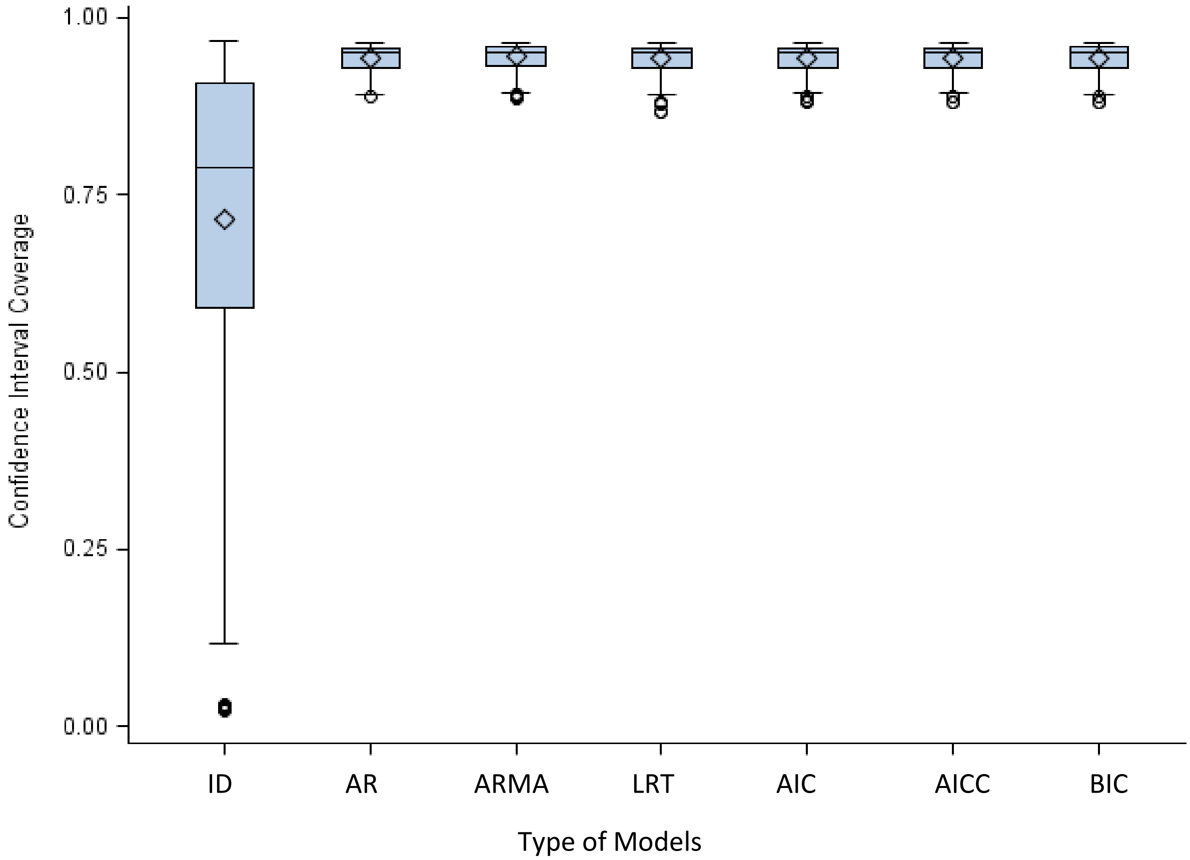

Figure 70. Box plots showing the distribution of the confidence interval coverage for the level-two phase effect (shift in level) across the seven models.

The box plot above illustrates the distribution of the confidence interval coverage for the level-two phase effect across the seven models. The smallest mean interval coverage was observed for the ID model (*M* = 0.71, *SD* = 0.25), conversely, the largest mean interval coverage was seen for the first order autoregressive moving average model (*M* = 0.94, *SD* = 0.19). To further examine the variability in the interval coverage, GLM models were run.

The results of the model, including three-way interactions, explained 99% of the total variability and resulted in the following factors being medium effects: the interaction between type of model and the variances of the error terms (η^2 = .08) and the interaction between the type of model and the level of the autocorrelation parameter (η^2 = .22).

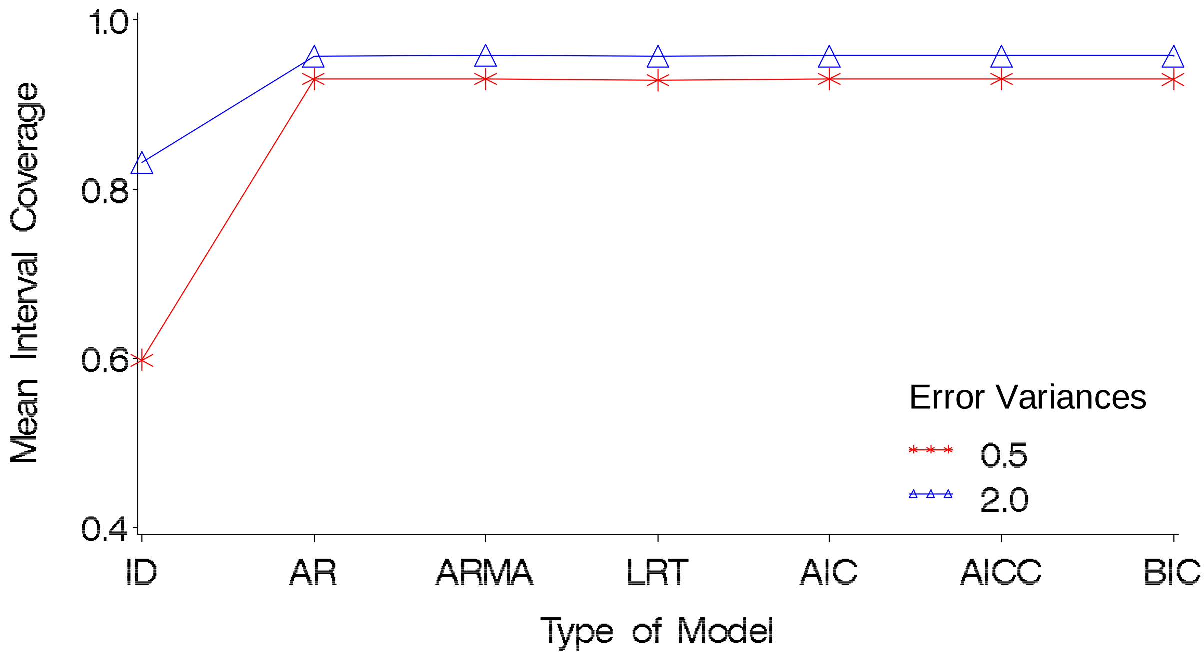

Figure 71. Line graphs illustrating the mean interval coverage for the level-two variance for the phase effect as a function of the interaction between the variances of the error terms and the type of model.

Line graphs were then created to further analyze the relationship of the mean interval coverage with these effects. The graph above (see Figure 71) displays that the association of the mean interval coverage with type of model depends on the variances of the error terms, specifically, the graph shows that the for the ID model when the most of the variance is at level-one, the mean interval width is much lower (*M* = 0.60, *SD* = 0.28) than for the other models (*M* = 0.93, *SD* = 0.017). However, this difference is much smaller for the ID model (*M* = 0.83, *SD* =

0.13) and the other models (M = 0.957, SD = 0.004) when most of the variance is at the upper levels.

Furthermore, the line graph (see Figure 72) below depicts the association of the level-two variance with the phase effect and the interaction between the level of the autocorrelation parameter and type of model. The graph below illustrates that for the ID model the effect on the mean interval coverage depends on the level of the autocorrelation parameter. Moreover, the figure illustrates that as the autocorrelation parameter is decreased from 0.4 (*M* = 0.52, *SD* = 0.26) to 0.2 (*M* = 0.79, *SD* = 0.12) to 0.0 (*M* = 0.95, *SD* = 0.01), then the mean interval coverage increases for the ID model. The mean interval coverage was near the nominal value of 0.95 for the six remaining models.

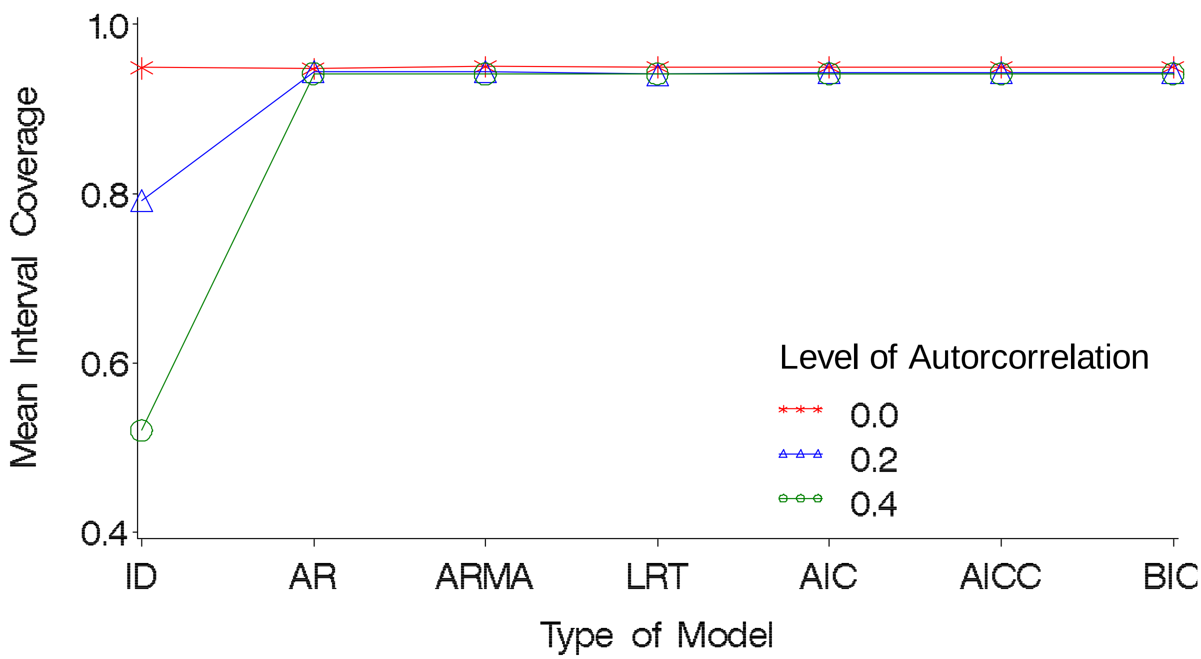

Figure 72. Line graphs representing the association between the mean interval coverage and the interaction between the level of the autocorrelation parameter and the type of model.

Level-two variance for the overall average treatment effect for the interaction (shift in slopes). The box plots illustrating the distribution of the confidence interval coverage for the level-two variance for the interaction effect (shift in slopes) is displayed in the Figure 73 below. The interval coverage seemed to be comparable across all models, with the exception of the ID model. The mean interval coverage for the ID model was 0.72 (*SD* = 0.24) which was lower than the mean interval coverage for the remainder of the models (*M* = 0.94, *SD* = 0.03). To further explore the variability of the interval coverage, GLM models were run. The results of the model (including three-way interactions) explained 99% of the total variability. The model resulted in one medium effect, the interaction between the level of the autocorrelation parameter and the type of model (η^2 = .22).

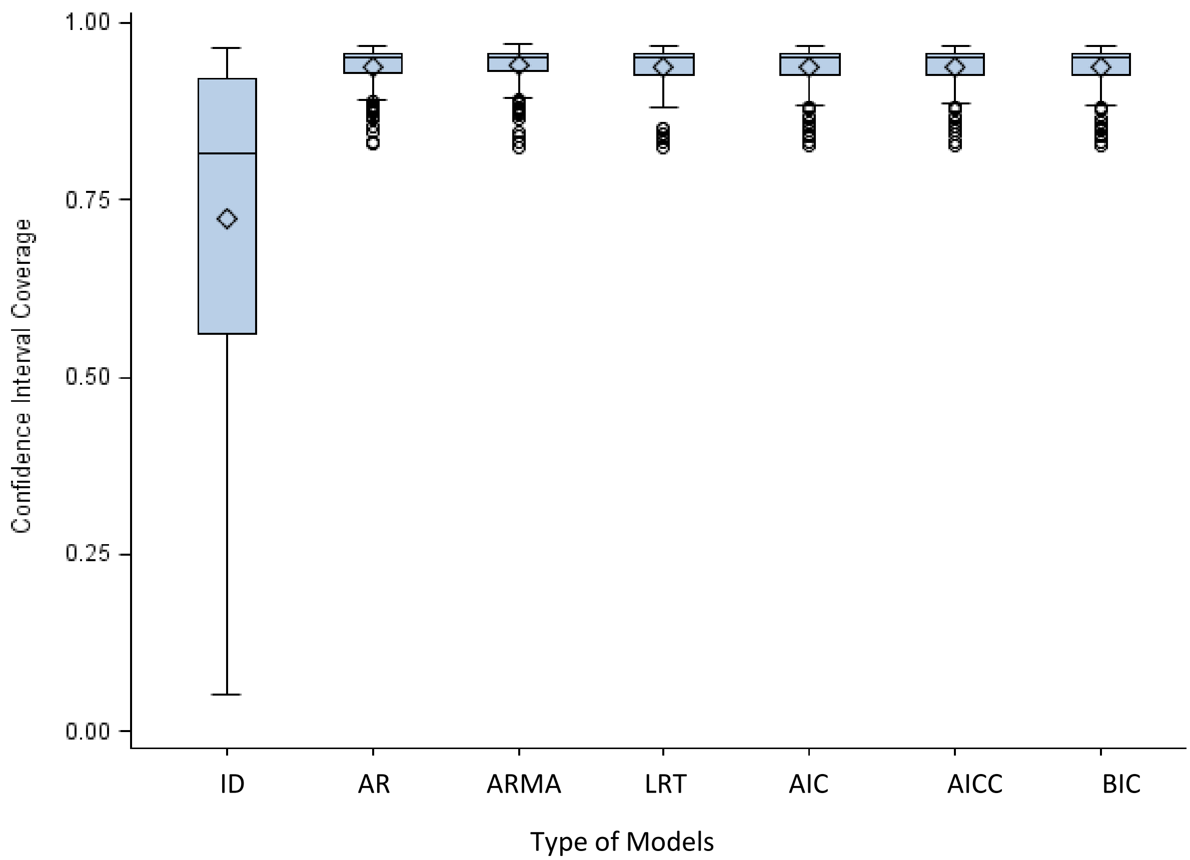

Figure 73. Box plots depicting the distribution of the confidence interval coverage for the level-two variance for the interaction effect (shift in slopes) across the seven models.

The graph (see Figure 74) was then constructed to further analyze the association between the mean interval coverage and the interaction between the level of the autocorrelation parameter and the type of model. The figure below represents that for the ID model, the effect on the mean interval coverage rely on the level of the autocorrelation parameter. Specifically, for the ID model the mean interval coverage is much lower when the autocorrelation parameter is 0.4 (*M* = 0.53, *SD* = 0.25) and then the mean interval coverage approaches the nominal value of 0.95 and the autocorrelation parameter approaches 0. The remaining six models all have interval coverage that was close to the nominal value of 0.95.

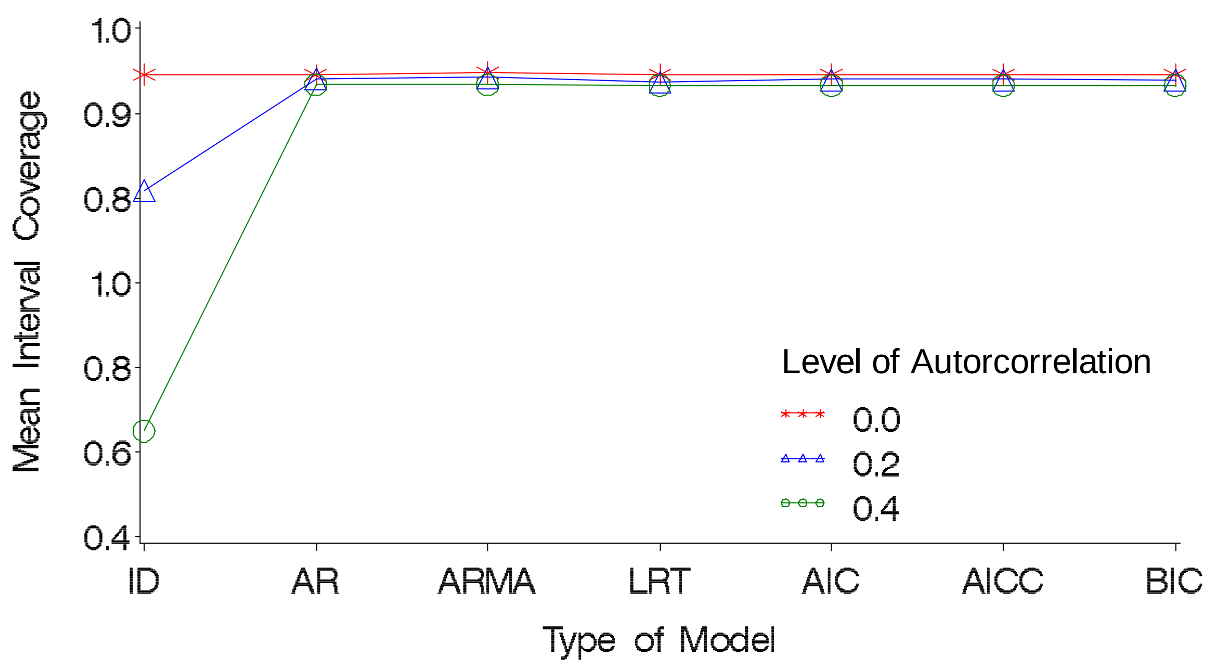

Figure 74. Line graph illustrating the relationship between the mean interval coverage and the interaction effect of the level of the autocorrelation parameter and the type of model.

Level-one or Residual Variance. The confidence interval coverage for the level-one variance or the residual variance was then analyzed. The smallest interval coverage (*M* = 0.56, *SD* = 0.32) was observed for the ID model, meanwhile the largest mean interval coverage was noted for the first order autoregressive moving average model, ARMA(1,1) (M = 0.70, SD = 0.27). The variability in the mean interval coverage was then explored using GLM models.

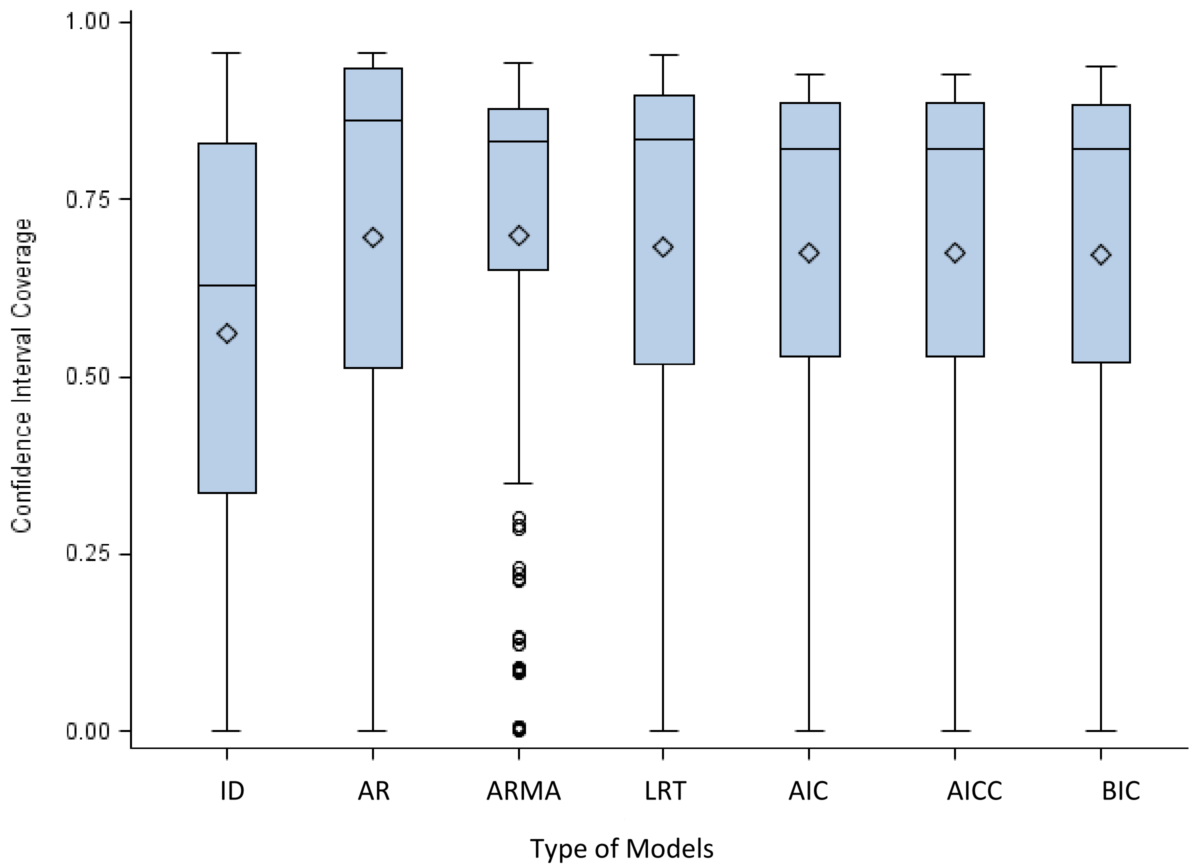

Figure 75. The distribution of the confidence interval coverage for the level-one variance across the seven models.

The models, including three-way interactions revealed that there were three medium effects: the number of primary studies to be included in the meta-analysis ($\eta^2 = 0.06$), the interaction of the series length and the type of model ($\eta^2 = 0.07$), and lastly, the level of the autocorrelation parameter ($\eta^2 = 0.53$). Box plots were then created to further examine the association of these effects with the mean interval coverage for the level-one variance (residual variance).

The graph below (see Figure 76) shows that as the number of primary studies to be included in the meta-analysis increased from 10 (M = 0.74, SD = 0.23) to 30 (M = 0.59, SD = 0.34), then the mean interval coverage for the level-one variance decreased.

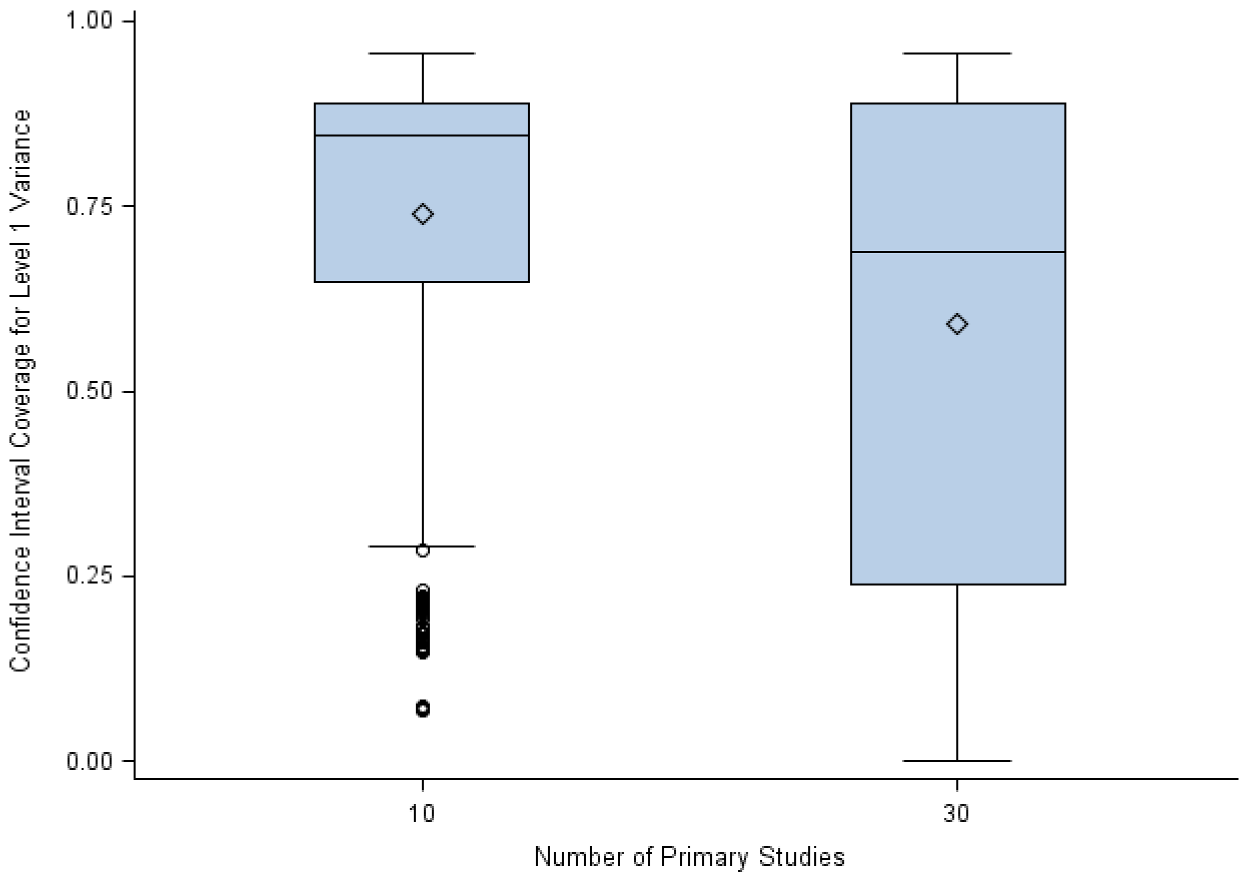

Figure 76. Box plots illustrating the relationship between the mean interval coverage and the number of primary studies to be included in the meta-analysis.

The graph in Figure 77 below displays the association of the mean interval coverage for the level-one variance with the interaction of the type of model and the series length. Specifically, the graph illustrates that the effect of the series length on the mean interval coverage depends on the type of model. The mean interval width is decreased for most of the models (AR, ARMA, LRT, AIC, AICC, BIC) when the series length is increased from 10 to 30. However, the opposite is observed for the ID model: the mean interval coverage is increased when the series length is increased from 10 (*M* = 0.43, *SD* = 0.35) to 30 (*M* = 0.69, *SD* = 0.21).

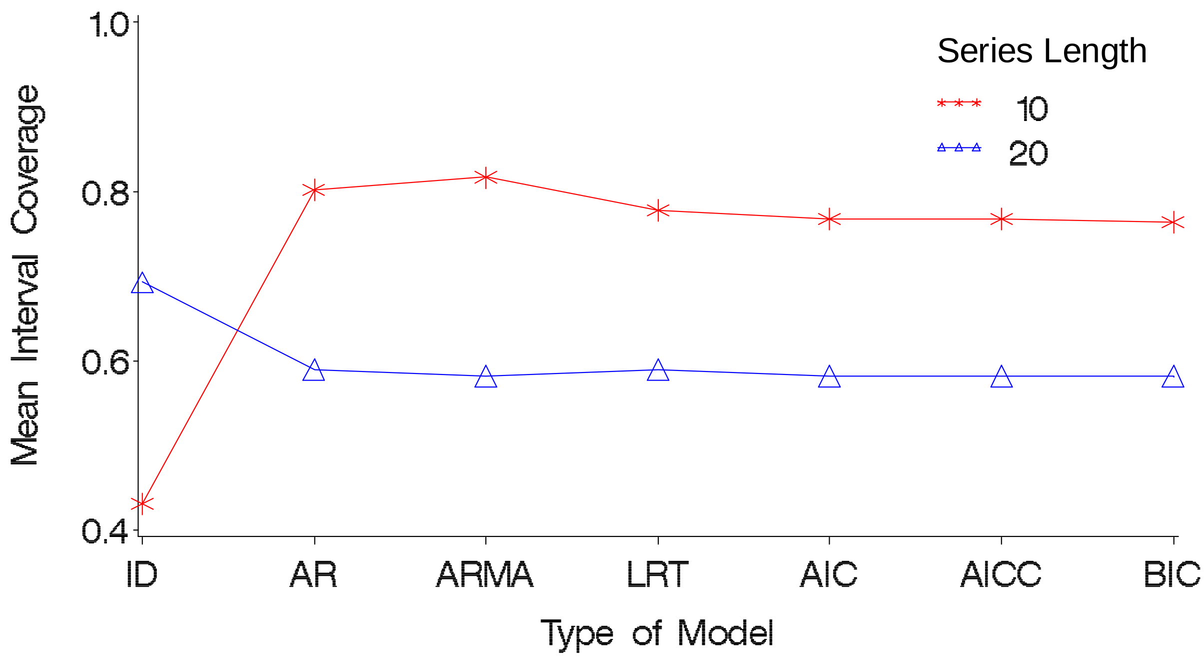

Figure 77. Line graphs depicting the relationship between mean interval coverage for the residual variance and the interaction of series length and the type of model.

Finally, the graph of the level of the autocorrelation parameter and the mean interval coverage is displayed below. The graph (see Figure 78) depicts that the mean interval coverage for the residual variance decreased as the level of the autocorrelation parameter increased from 0.0 (*M* = 0.91, *SD* = 0.06) to 0.2 (*M* = 0.81, *SD* = 0.14) to 0.4 (*M* = 0.40, *SD* = 0.29). The amount of variance for the interval coverage for the residual variance also increased as the level of autocorrelation increased.

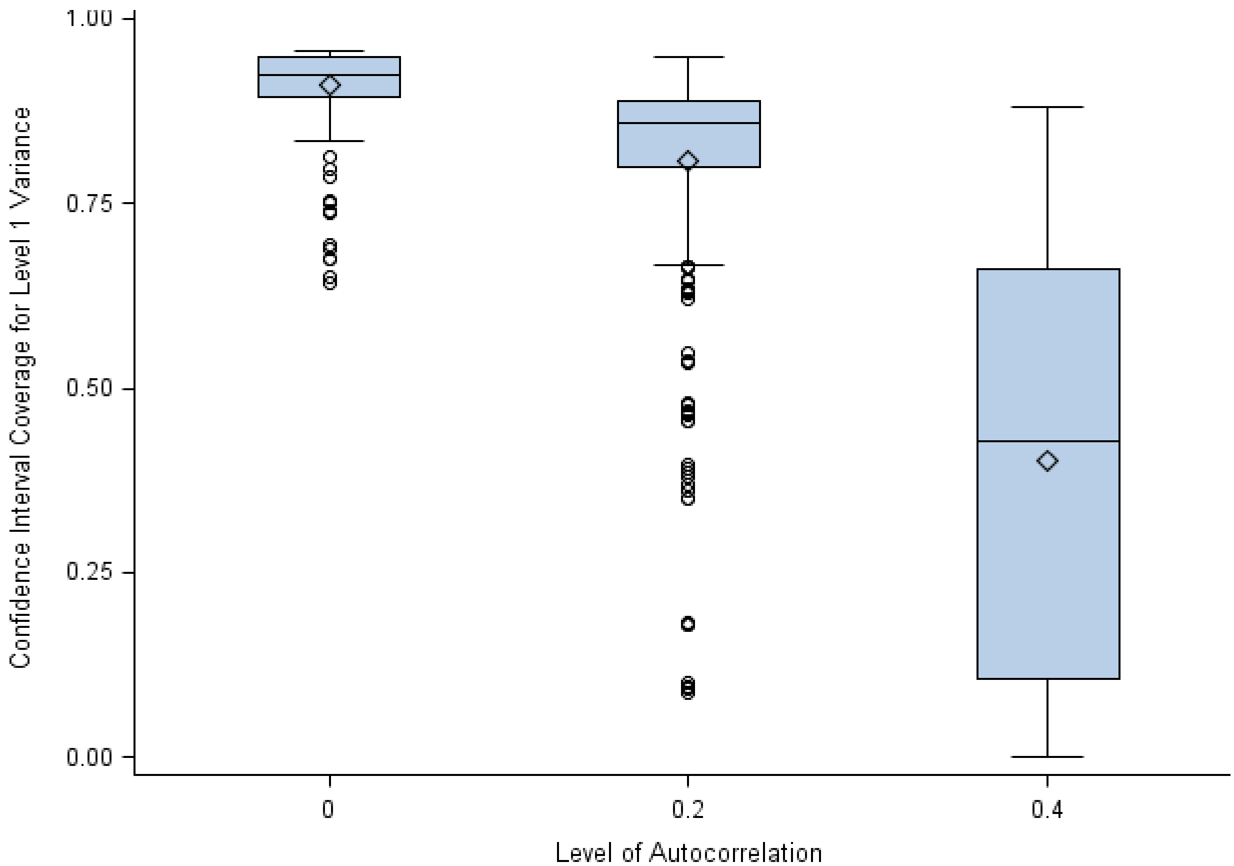

Figure 78. Box plots depicting the relationship between the mean interval coverage for the residual variance and the level of the autocorrelation parameter.

Autocorrelation parameter. Box plots (see Figure 79) illustrating the distribution of the confidence interval coverage for the autocorrelation parameter across the six models (the autocorrelation parameter was estimated as 0 for the ID model).. The smallest mean interval coverage was noted for the model selected by the AIC (M = 0.86, SD = 0.06) and the largest mean interval coverage was for the AR(1) model (M = 0.95, SD = 0.005). The variability was further examined with the use of GLM models. The model, including three-way interactions explained 96% of the total variability. The model also revealed one significant effect: the interaction of the level of the autocorrelation parameter and the type of model (η^2 = 0.32). Line graphs were then created to further examine the association of the mean interval coverage with this effect.

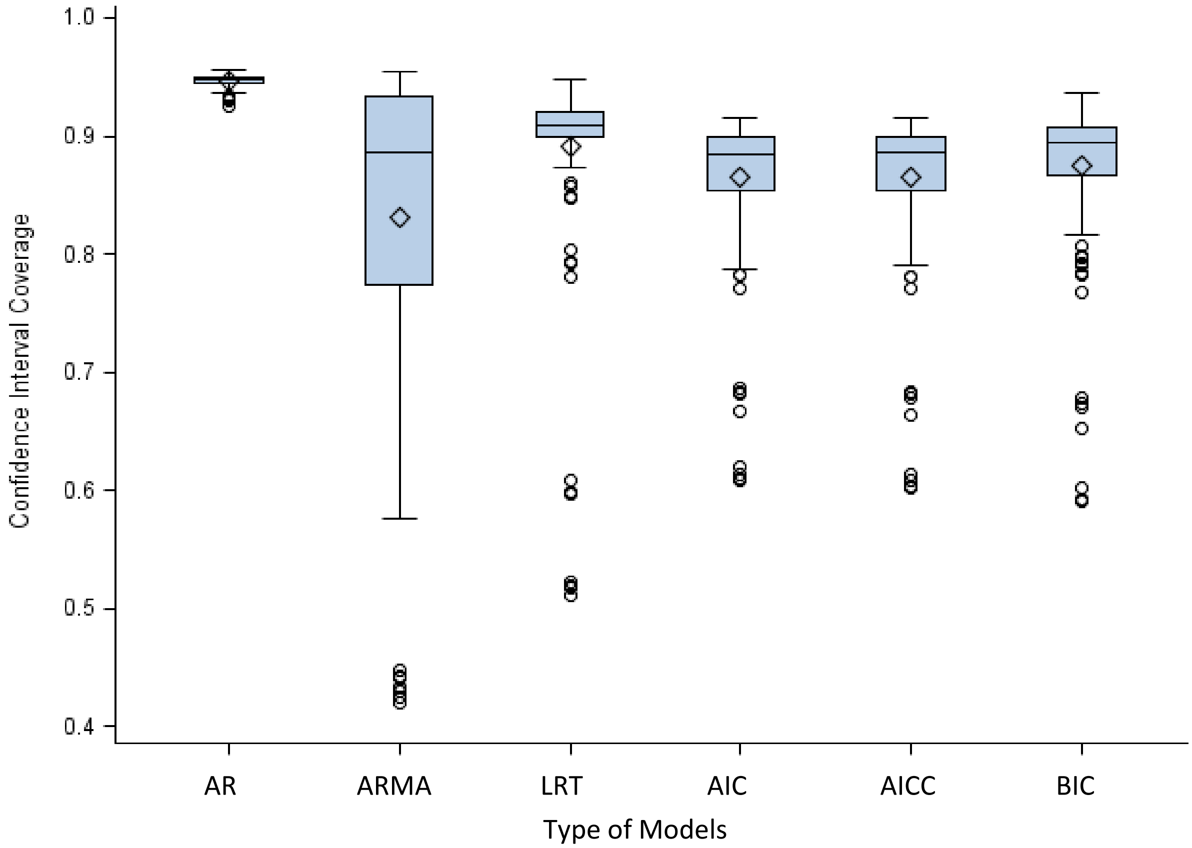

Figure 79. Box plot displaying distribution of the confidence interval coverage for the autocorrelation parameter across the six models.

The association between mean interval coverage for the autocorrelation parameter and the interaction of the level of the autocorrelation parameter and the type of model is depicted in the line graph below (see Figure 80). The graph indicates that the relationship across the models selected by the fit indices are comparable, that the mean interval coverage is greatest when ρ = 0.0 and least when ρ = 0.2. The last model, the ARMA (1,1) model, illustrates that the mean interval coverage increases as the level of the autocorrelation parameter increases from 0.0 (M = 0.60, SD = 0.12) to 0.2 (M = 0.86 ,SD = 0.07) to 0.4 (M = 0.92, SD = 0.04).

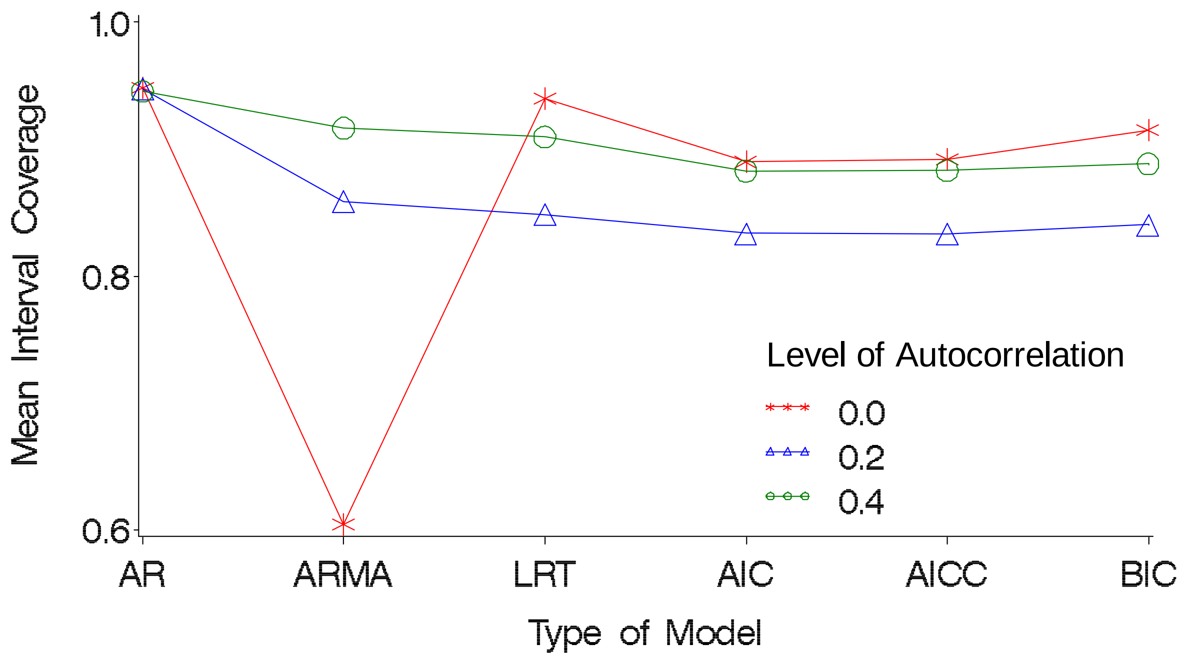

Figure 80. Line graph depicting the association between mean interval coverage for the autocorrelation parameter and the interaction of the level of autocorrelation parameter and the type of model.

First-Order Autoregressive Moving Average Parameter. The mean confidence interval coverage was then examined for the moving average parameter across the five models (the parameter was estimated to be zero for both the ID and the AR models) in Figure 81. The largest mean interval coverage was observed for the models selected by LRT (M = 0.59, SD = 0.45), conversely, the smallest mean interval coverage was noted for the ARMA(1,1) model (M = 0.54, SD = 0.41). The variability was further explored using GLM models. The model, including two-way interactions, explained over 99% of the total variability, and revealed two medium effects: the level of the autocorrelation parameter ($\eta^2 = 0.18$) and the interaction between the level of the moving average parameter and the type of model ($\eta^2 = 0.54$).

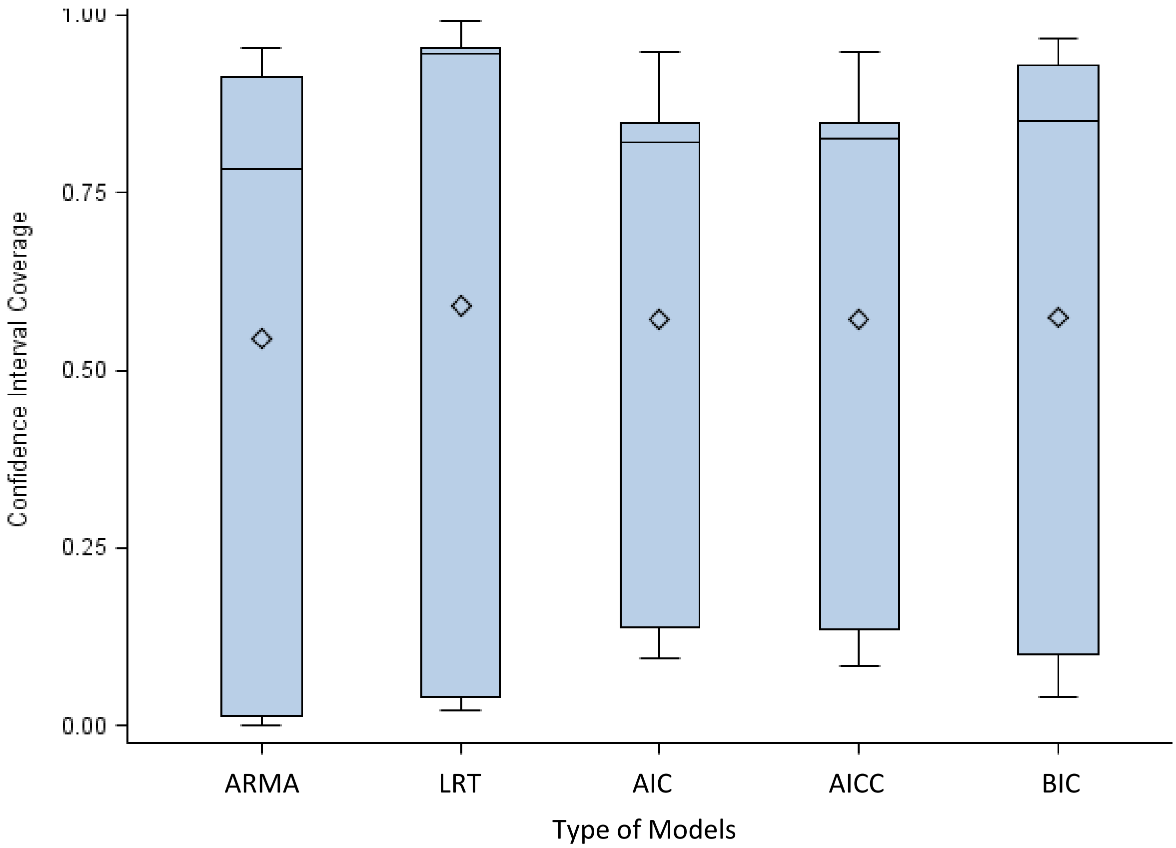

Figure 81. Box plots illustrating the distribution of the confidence interval coverage for the moving average parameter across the five models.

Graphs were then constructed to analyze the means of the interval coverage for the moving average parameter across both of the significant effects. First, the graph (see Figure 82) shows that the mean coverage for the moving average parameter decreases as the level of autocorrelation increases from 0.0 (M = 0.91, SD = 0.08) to 0.2 (M = 0.49, SD = 0.40) to 0.4 (M = 0.48, SD = 0.41). The variability is increased greatly as the autocorrelation increases for 0.0 to 0.2 but remains the approximately the same as the autocorrelation is further increased from 0.2 to 0.4.

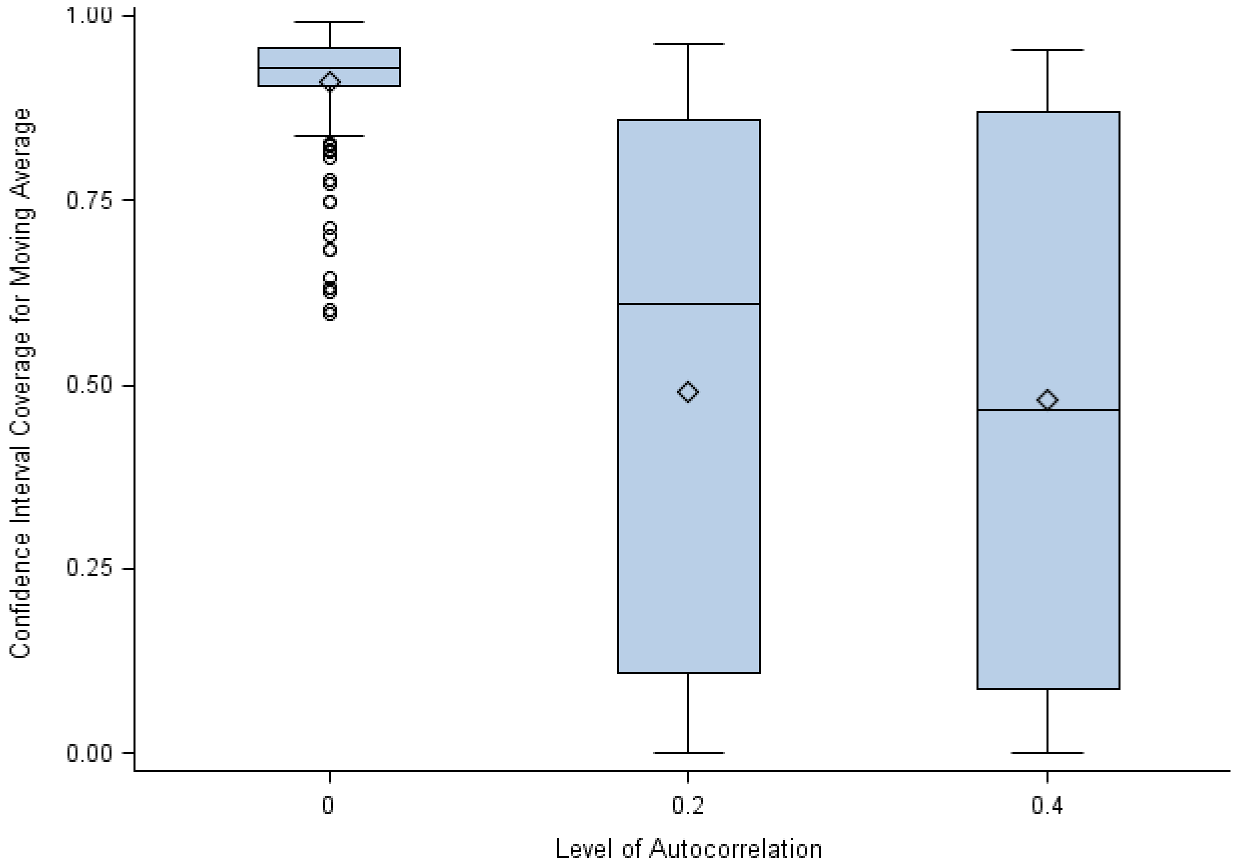

Figure 82. Box plots illustrating the mean interval coverage for the moving average parameter across the levels of the autocorrelation parameter.

Additionally, the association of the mean interval coverage for the moving average parameter and the interaction of the level of the moving average parameter and the type of model are depicted in the line graph below (see Figure 83). Analyzing the graph below, there is little difference in the mean coverage for the moving average parameter across all of the five models as the level of the moving average parameter is increased from 0.2 to 0.4. Additionally, the mean coverage is really low for the models selected by the fit indices (this again can be attributed to the few number of times that the fit indices correctly identified the first order autoregressive moving average model). However, for the ARMA (1,1) model, the mean coverage was high when the moving average parameter was at least 0.2 (M = 0.89, SD = 0.06). Lastly, when the moving average parameter was 0 (ID model), then the mean coverage was lowest for the ARMA (1,1) model (M = 0.30, SD = 0.36) and highest for the LRT selected models (M = 0.96, SD = 0.02).

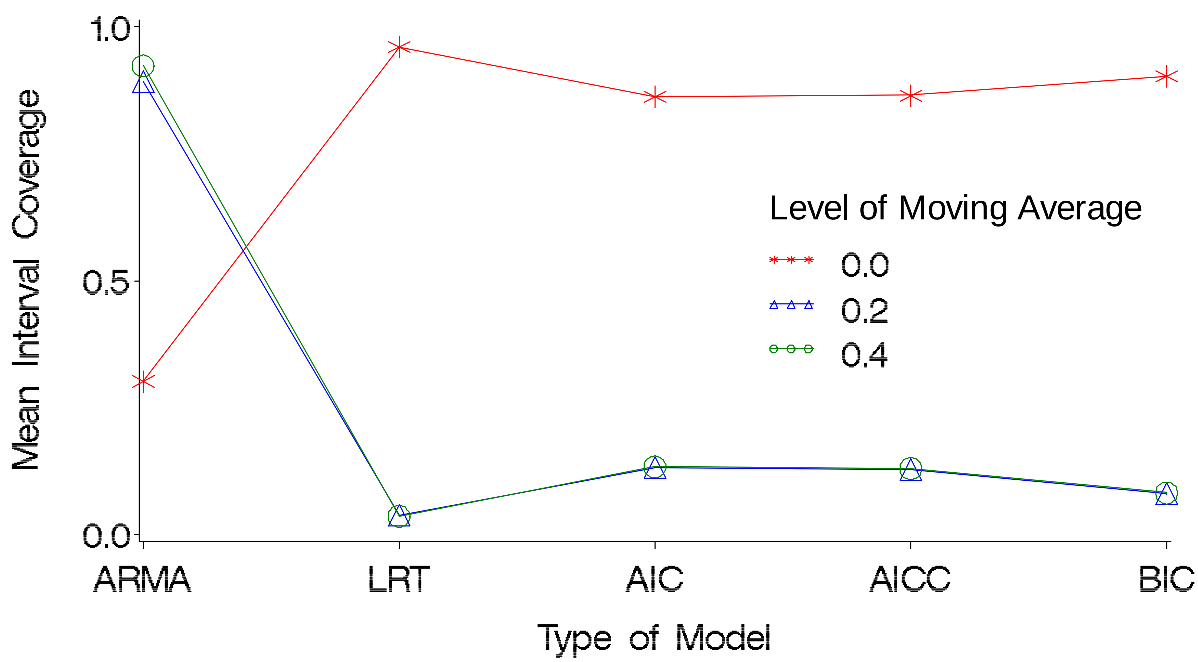

Figure 83. Line graphs depicting the association of the mean interval coverage for the moving average parameter and the interaction of the type of model and the level of the moving average parameter.

Confidence Interval Width

The interval width for the four level-three and level-two variance components were so large, that they provided no valuable information. Specifically, the smallest mean interval width estimates for the level-three variance components for the phase effect, shift in level, was for the ID model, 7.08 X 10^{283} and largest for the model selected by the BIC, 4.76 X 10^{284}. For the level-three variance components for the level-three interaction effect (shifts in slopes), the smallest mean interval width was observed for the AR model, 5.83 X 10^{283} and smallest mean interval width was noted for the model selected by the AICC, 4.76 X 10^{284}. Similar patterns were observed for the level-two variance components, for the phase effect, the smallest mean interval width was seen for the ID model, 6.89 X $10^{283}$and the largest mean interval width was for the model selected by LRT, 1.97 X 10^{284} . Finally, when analyzing the means for the level-two variance for the interaction effect, the shift in slopes, and the smallest mean interval width was noted again for the ID model, 1.17 X 10^{282} and the largest mean interval width was observed for the ARMA model, 1.53 X 10^{284}.

Level-one or Residual Variance. The distribution for the confidence interval width for the level-one variance for two of the models (for ID and for the AR model); the remaining models had widths that were too large to gain any meaningful information for the level-one variance (residual variance). Therefore, the other models were removed from the picture to allow for an accurate examination of the ID and the AR model.

The mean interval width for the ID model (*M* = 0.17, *SD* = 0.07) was smaller than for the AR model (*M* = 0.34, *SD* = 0.25) for the level-one variance. GLM models were run to further examine the variability in the interval width for the residual variance across the design factors and the combination of these design factors. The model, including three-way interactions, explained 95% of the total variability, and revealed that there were four medium effects: the number of participants per study (η^2 = 0.06), the interaction of the autocorrelation parameter and the type of model (η^2 = 0.08), the number of primary studies to be included in the meta-analysis (η^2 = 0.14), and the series length or number of observations (η^2 = 0.15). Graphs were then created to further examine these effects and their relationship with the outcome of interest (mean interval width).

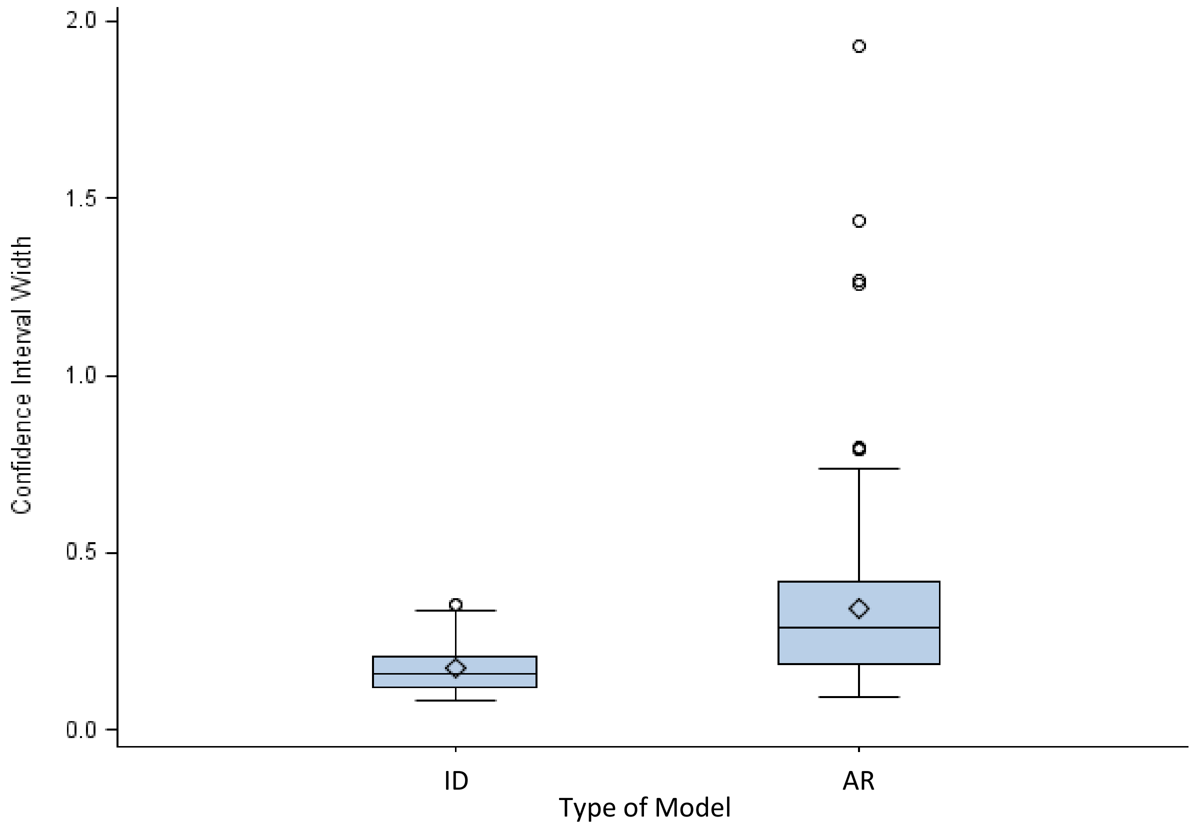

Figure 84. Box plot illustrating the distribution for the confidence interval width for the level-one residual variance across the seven models.

Figure 85 below displays the relationship between the mean interval width for the level-one variance and the number of participants in a study. The graph shows that as the number of participants increased in a study from 4 (M = 0.31, SD = 0.24) to 8 (M=0.21, SD = 0.13), then the mean interval width for the level-one variance also increased. In terms of the variability, the box plot also revealed that the variance tended to tended to decrease as the number of participants to be included in each study increased from 4 to 8.

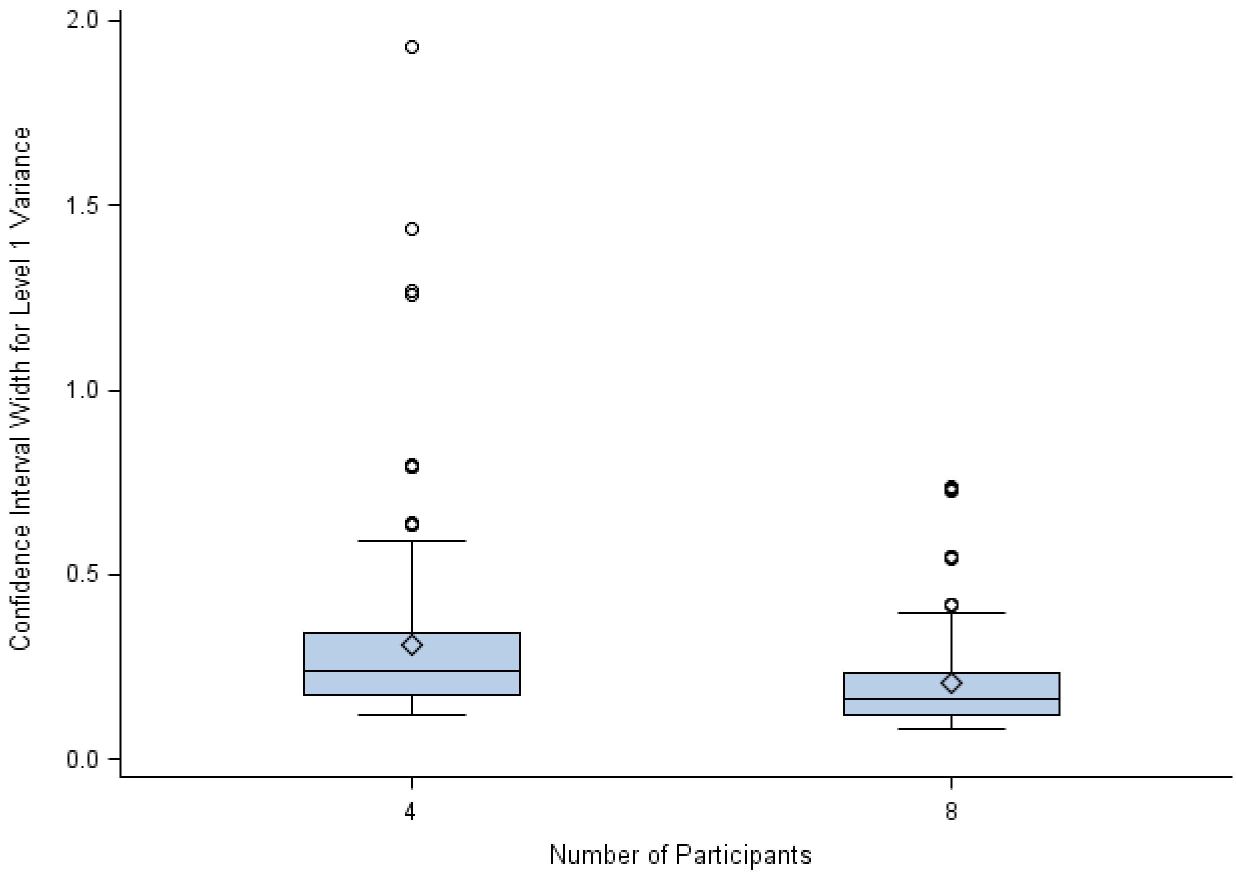

Figure 85. Box plots illustrating the association of the mean interval width for the level-one variance and the number of participants in a particular study.

Next, the relationship between the mean interval width and the interaction of the level of the autocorrelation parameter and the type of model is displayed in Figure 86 below. The graph shows that the effect of the level of the autocorrelation parameter on the mean interval width depends on the type of model. Specifically, when the type of model is ID, the level of autocorrelation has very little effect on the mean interval width for the level-one variance. However, for the AR (1) model, the mean confidence interval width for the level-one variance increases as the level of the autocorrelation parameter increases from the model not having any autocorrelation to the highest level of autocorrelation.

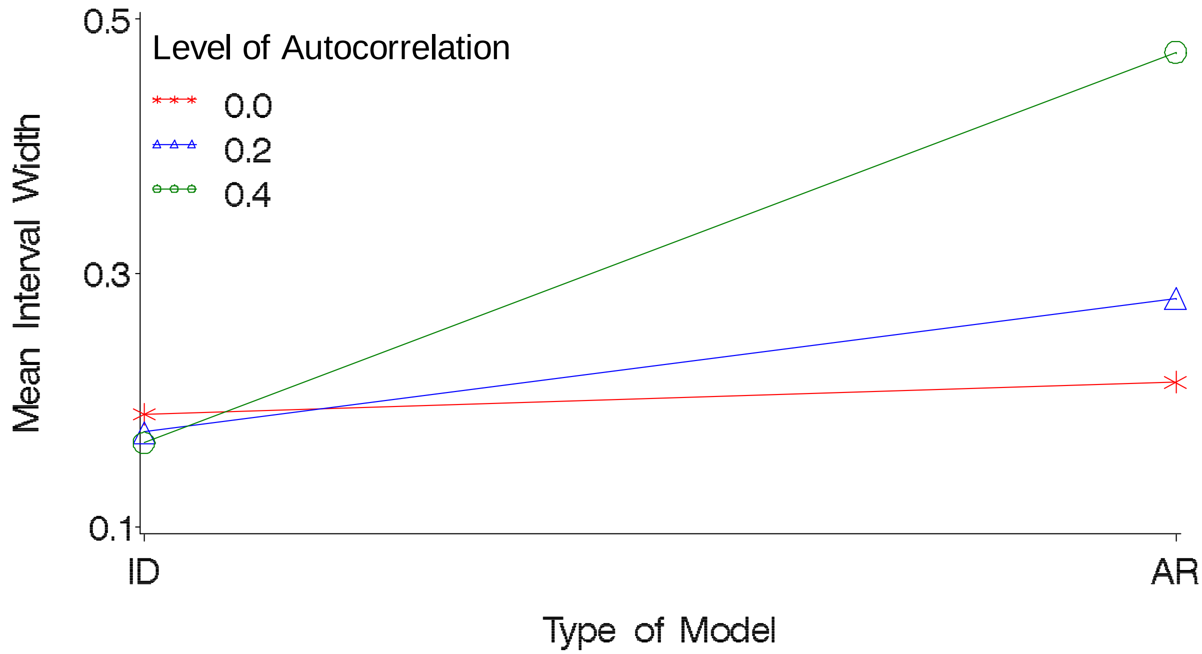

Figure 86. Line graph displaying the relationship of the mean interval width for the level-one variance and the interaction of the level of the autocorrelation parameter and the type of model.

Additionally, the association of the mean interval width for the residual variance with the number of studies to be included in the meta-analysis was further analyzed using graphs (see Figure 87 below). The graph indicates that as the number of primary studies to be included in the meta-analysis increased from 10 (*M* = 0.33, *SD* = 0.24) to 30 (*M* = 0.18, *SD* = 0.11), then the mean interval width for the residual variance also decreased. The variability also tended to decrease as the number of primary studies increased from 10 to 30.

Lastly, the association of the mean interval width for the residual variance and the number of observations or the series length was analyzed. The graph in Figure 88 below illustrated that as the series length increased from 10 (*M* = 0.34, *SD* = 0.18) to 20 (*M* = 0.18, *SD* = 0.08), then the mean interval width for the residual variance decreased. The descriptive statistics also revealed that the variability tended to decrease with the increased series length.

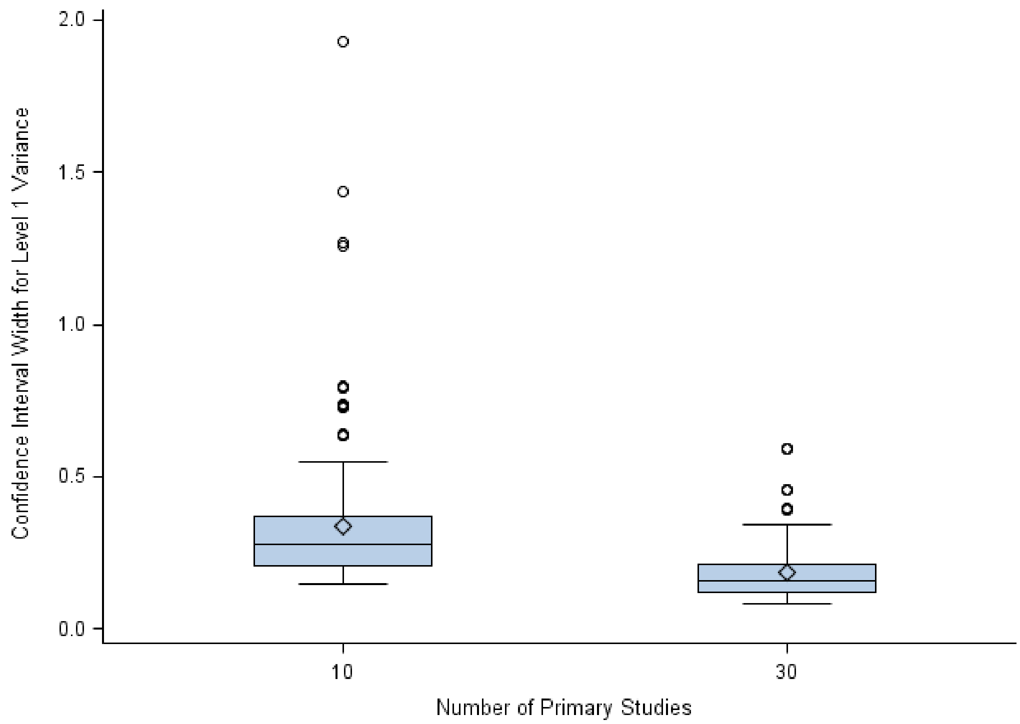

Figure 87. Box plots depicting the relationship between the mean interval width for the level-one variance and the number of primary studies included in the meta-analysis.

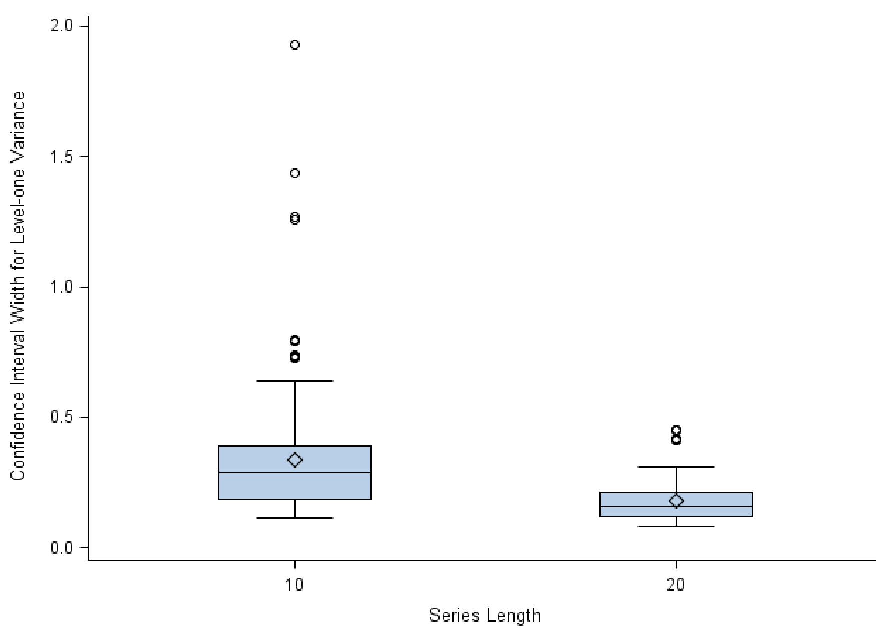

Figure 88. Box plot illustrating the relationship between the mean interval width for the level-one variance and the series length.

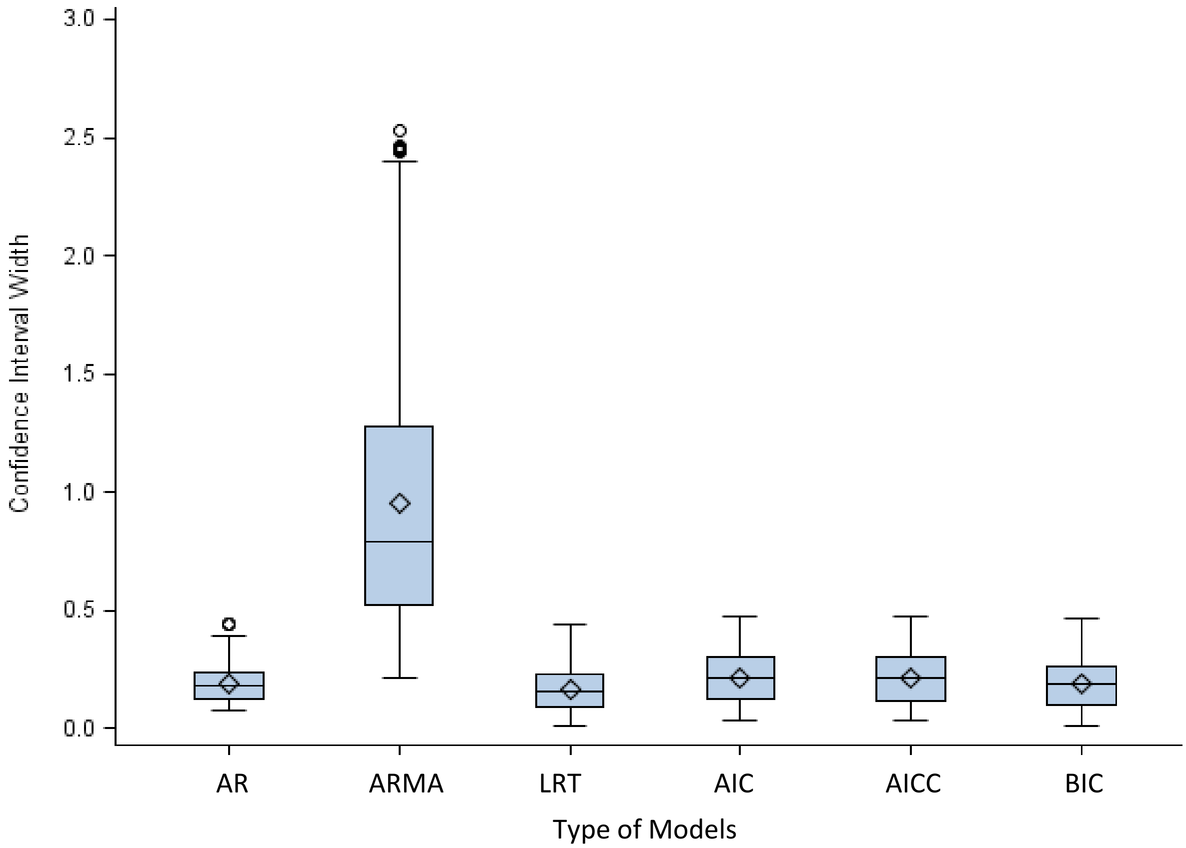

Figure 89. The box plot representing the distribution for the interval width for the autocorrelation parameter across the six models.

Autocorrelation Parameter. The distribution for the interval width for the autocorrelation parameter is shown above in Figure 89. The box plot shows that the means for the interval width are different among the models, the ARMA (1, 1) model has the largest mean interval width (M = 0.96, SD = 0.59) and the model selected by the LRT has the smallest mean interval width (M = 0.16, SD = 0.10). To further explore the variability in the mean interval width, GLM models were run across the design factors.

The model, including two-way interactions, explained 98% of the total variability and resulted in one medium effect, the interaction of autocorrelation parameter and the type of model

(η^2 = 0.31). Line graphs were then constructed to further examine the relationship of the mean interval width and the interaction effect. The line graph below displays that the effect of the level of the autocorrelation parameter depend on the type of model. Specifically, for the first order autoregressive moving average model, the mean interval width for the autocorrelation parameter is much greater than for the remainder of the models.

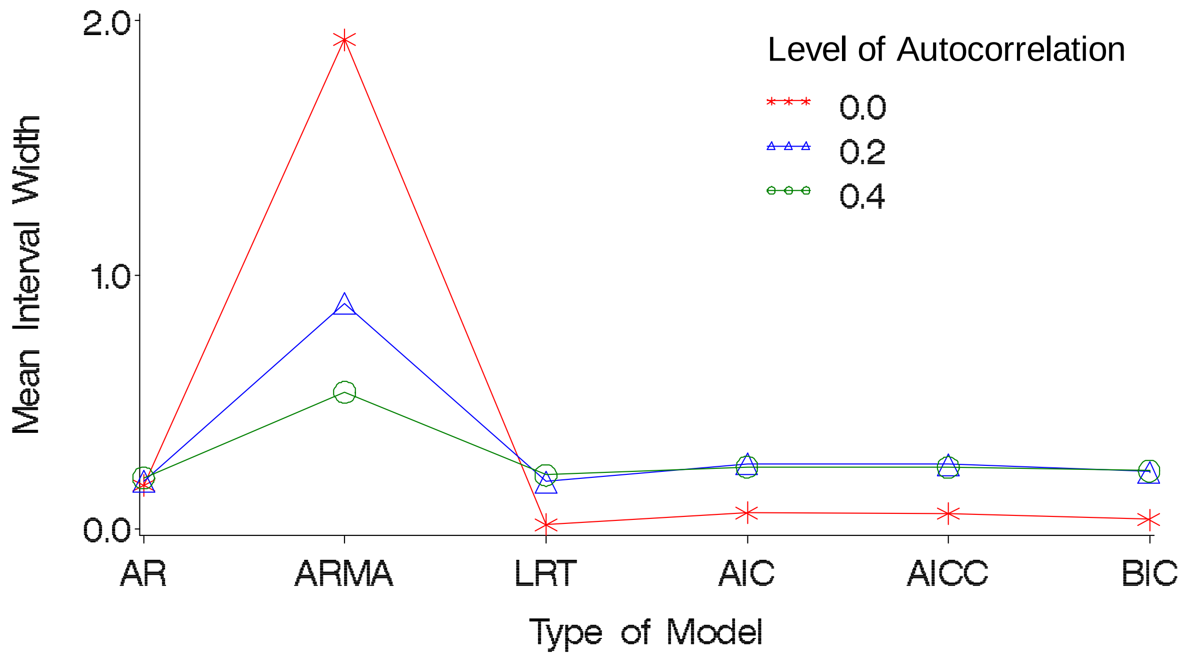

Figure 90. Line graph depicting the estimated mean interval width of the autocorrelation parameter as a function of the interaction between the level of the autocorrelation parameter and the type of model.

Moreover, the mean interval width decreases for the ARMA (1,1) model when the level of the autocorrelation parameter increases from 0.0 (*M* = 1.92, *SD* = 0.39) to 2.0 (*M* = 0.89, *SD* = 0.31) to 4.0 (*M* = 0.54. *SD* = 0.24). For the autocorrelation model, the mean interval width slightly increases as the level of the autocorrelation parameter increases from 0.0 (*M*= 0.17, *SD* = 0.08) to 4.0 (*M* = 0.20, *SD* = 0.10). When looking at the fit-selected models, overall the mean interval width is comparable, resulting in mean interval widths of less than 0.10 when the level of the autocorrelation parameter was 0.0 and increased to approximately 0.24 when the autocorrelation parameter increased to 0.2 or 0.4

First-Order Autoregressive Moving Average Parameter. The distribution for the mean confidence interval width for the moving average parameter across the five models are displayed in Figure 89 below. The box plots illustrate that the means for the four models selected by fit indices were comparable with the largest mean for the model selected by the AIC (*M* = 0.04, *SD* = 0.02) and smallest for the model selected by the LRT (*M* = 0.01, *SD* = 0.01). However, the first-order autoregressive moving average model was the greatest among the five models with a mean of 8.30 (*SD* = 20.00). To further examine the variability, GLM models were run.

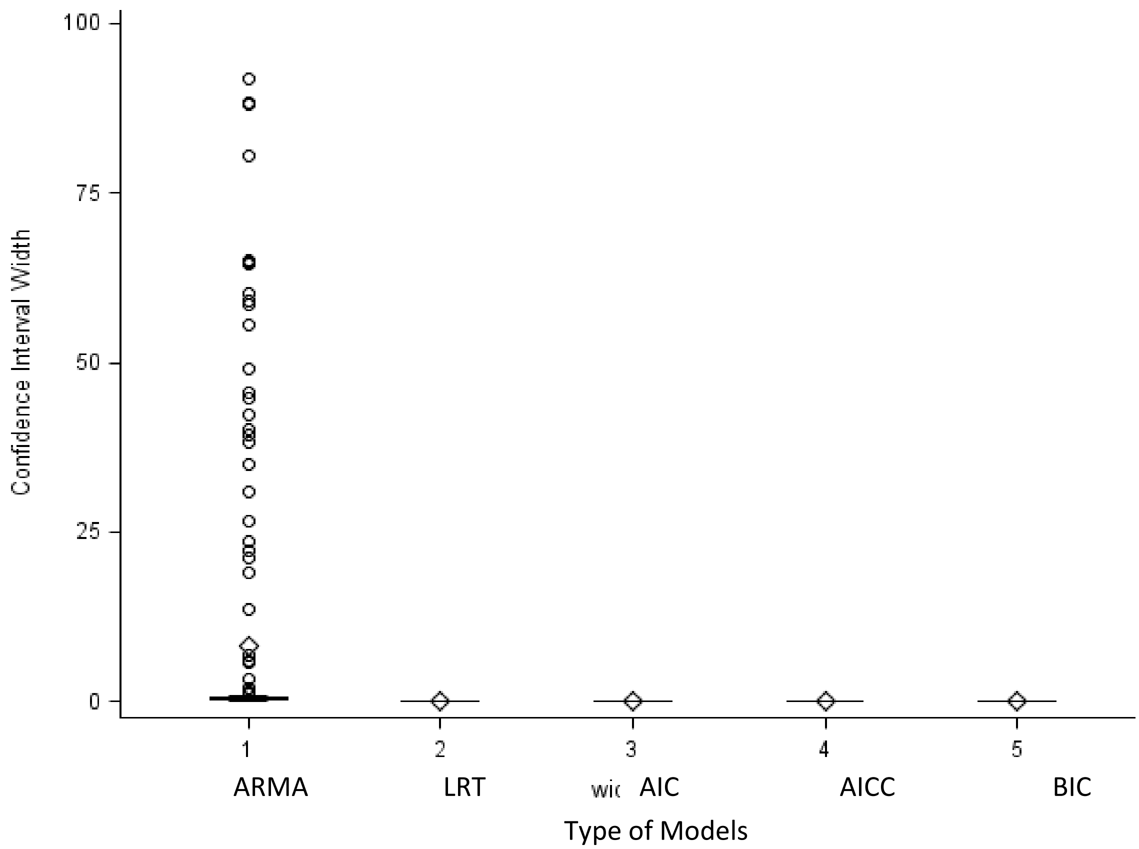

Figure 91. Box plots illustrating the distribution of the confidence interval width estimates for the moving average parameter across the five models.

The resulting model, including five-way interactions, explained 96% of the total variability and resulted in one medium effect, the interaction of the level of the autocorrelation parameter and type of model ($\eta^2 = 0.28$). Line graphs were then created to further examine the association of the mean interval width for the moving average parameter with this effect.

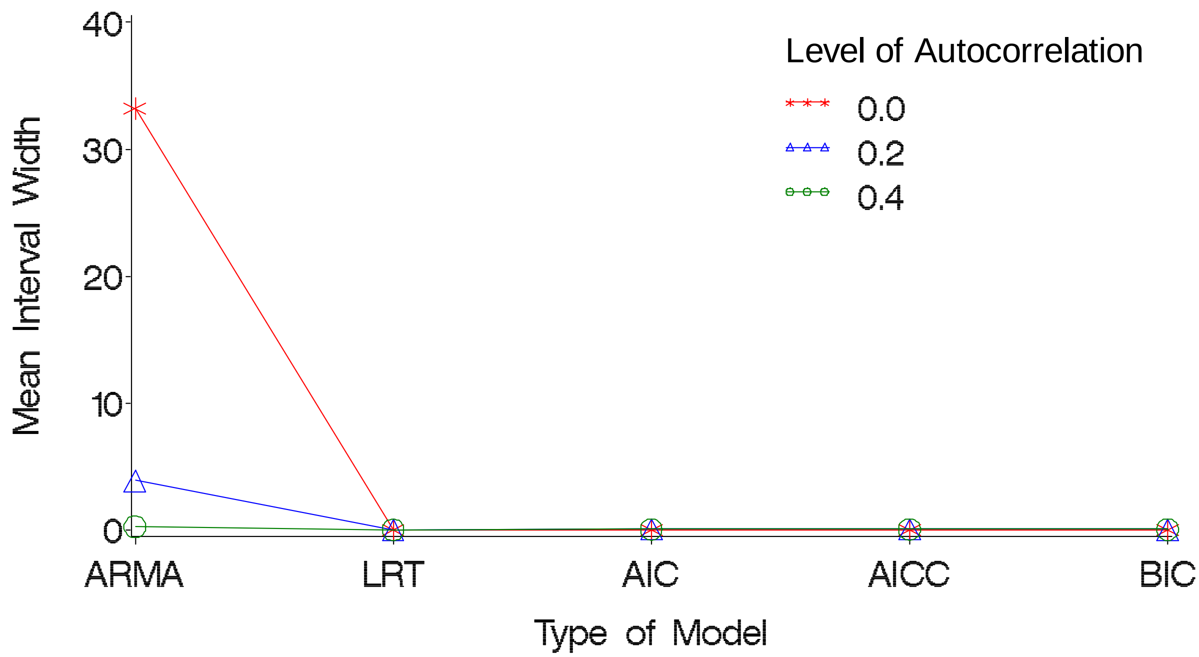

Figure 92. Line graphs illustrating the association of mean interval width for the moving parameter and the interaction of the level of the autocorrelation parameter and the type of model.

The line graph above (see Figure 92) demonstrates that the models selected by the fit indices had comparable mean interval widths for the moving average parameter across all levels of the autocorrelation parameter (this should be interpreted with caution again due to the few number of times that these models correctly identified the ARMA model). However, for the ARMA(1,1) model, there were great differences in the mean interval widths as the level of the autocorrelation parameter increased from 0.0 (*M* = 33.22, *SD* = 27.18) to 0.4 (*M* = 0.25, *SD* = 0.13).

The distribution for the proportion of times that the ARMA (1,1) model was correctly identified was examined. The results revealed that none of the fit indices correctly selected the model more than 20% of the times. The results also indicated that the greatest mean proportion of times was for the AIC, meanwhile the LRT had the smallest proportion of correct identification for this model. The association of the mean proportion of times that the ARMA model was correctly identified by the fit indices and the interaction effect of the type of fit index and the number of primary studies to be included in the meta-analysis was then explored. The results indicated that the effect of the number of primary studies on the mean proportion of correct identification depended on the type of model that was used. Concretely, for the BIC fit index, the greatest decrease in the correct identification was seen when the number of primary studies increased. However, there was hardly any difference for the models selected by the LRT in the mean proportion of correct selection when the number of primary studies increased.

Table 6

Summary of Results for Correct Model Selection

ID	AR(1)	ARMA(1,1)
M = 0.74; Range: [0.82, 0.91].	*M* = 0.84; Range: [0.52, 0.95].	*M* = 0.14; Range: [0.05, 0.24].
The models selected by the BIC fit index showed the greatest improvement when the number of primary studies included in meta-analysis increased.	The proportion of times that the AR (1) model was correctly identified increased with increased series length and the number of primary studies to be included in the meta-analysis increased.	None of the fit indices correctly selected the model more than 25% of the time; the greatest proportion of times was found for the AIC fit index, while the least was for the LRT.
	The LRT fit index greatly outperformed the other fit indices in terms of correct identification for the simpler models, ID and AR(1).	For the BIC fit index, there was an increase in correct identification when the number of studies to be included in the meta-analysis increased.

Table 7
Summary of Results for the Fixed Effects

Parameter Estimate	Bias	RMSE	Interval Coverage	Interval Width	Type I Error	Power
Shift in Level	*M* = 0.00001;Range: [-0.0147, -0.0140]. No medium or larger effects were found.	*M* = 0.32 Range: [0.15, 0.55] Tended to decrease with increased number of primary studies and as most of the variance shifted to most of the variance at level-one.	*M* = 0.95; Range: [0.94, 0.97]. Tended to approach the nominal value of 0.95 across all design factors.	*M* = 1.52; Range: [0.60, 22.17]. Tended to decrease with increased level-three sample size and most of the variance at level-one. For the ARMA(1,1) model, the width becomes smaller as the level of autocorrelation parameter increased.	*M* = 0.05; Range: [0.04, 0.06]. Approached the target value across the models (M = 0.05) for all of the combinations of design factors.	*M* = 0.98; Range: [0.89, 1.0]. Mean Power estimates were comparable across the models (M = 0.98). When most of the variance is at the upper levels, mean power estimates increase with an increase in level-three sample size.

Table 7 (continued)
Summary of Results for the Fixed Effects

Parameter Estimate	Bias	RMSE	Interval Coverage	Interval Width	Type I Error	Power
Shift in Slope	*M* = 8.304X10^{-6}; Range: [-0.006, 0.006]. No medium or larger effects were found.	*M* = 0.10; Range: [0.04, 0.19]. Tended to decrease with increased number of primary studies and as most of the variance shifted to most of the variance at level 1.	*M* = 0.95; Range: [0.94, 0.96]. Tended to approach the nominal value of 0.95 across all design factors.	*M* = 0.49; Range: [0.18, 8.38]. Tended to decrease with increased level-three sample size and most of the variance at level-one. For the ARMA(1,1) model, the width becomes smaller as the level of autocorrelation parameter increased.	*M* = 0.05; Range: [0.04, 0.06]. The mean type I error approached the target value across the models (M = 0.05) for all of the combinations of design factors.	*M* = 0.53; Range: [0.15, 0.99]. The mean power estimate (M = 0.53) was approximately equivalent across the models. Tended to increase as the number of primary studies increased and most of the variance is at level-one.

Table 8
Summary of Results for Variance Components

	Relative Bias	RMSE	Interval Coverage	Interval Width
<u>Level-three</u>				
Shift in Level	*M* = 2.21, Range: [-0.023, 228.76] No medium or larger effects were found.	*M* = 101.87, Range: [0.172, 6183.13]. No medium or larger effects were found.	*M* = 0.96, Range: [0.91, 0.97]. Tended to overcover when most of the variance is at upper levels, but approached nominal value with increased level-three sample size. The impact of the level-two sample size was minimal when most of the variance is at the upper levels.	*M* = $3.72\text{x}10^{284}$, Range: [0.8616 to $4.03\text{x}10^{286}$]. No medium or larger effects were found.

Table 8 (Continued)
Summary of Results for Variance Components

	Relative Bias	RMSE	Interval Coverage	Interval Width
Shift in Slope	*M* = 3.23, Range: [-0.020, 368.73]. No medium or larger effects were discovered.	*M* = 9.243, Range:[0.0157, 400.754] No medium or larger effects were found.	*M* = 0.95, Range: [0.85, 0.98]. Tended to overcover, however this was magnified in the cases where the most of the variances was at the upper levels and series length was long.	*M* = $1.68x10^{284}$, Range:[0.0755, $4.03x10^{286}$]. No medium or larger effects were found.

Table 8 (Continued)
Summary of Results for Variance Components

	Relative Bias	RMSE	Interval Coverage	Interval Width
<u>Level-two</u>				
Shift in Level	*M* = 3.61, Range: [-0.03, 258.29]. No medium or larger effects were found.	*M* = 211.98, Range: [0.12, 14034.3]. No medium or larger effects were found.	*M* = 0.91, Range: [0.02, 0.97]. The coverage was close to 0 for the ID model for some instances when there was at least a moderate amount of autocorrelation. The level of autocorrelation did not have as great of an impact on the other models.	*M* = $1.64X10^{284}$, Range:[0.5204, $1.24x10^{286}$] No medium or larger effects were found.

Table 8 (Continued)
Summary of Results for Variance Components

	Relative Bias	RMSE	Interval Coverage	Interval Width
Shift in Slope	*M* = 4.87, Range: [-0.08, 419.24]. No medium or larger effects were found.	*M* = 9.24,Range: [0.0157, 400.75]. No medium or larger effects were found.	*M* = 0.91, Range: [0.05, 0.97]. The coverage was very low, close to 0, for the ID model when the level of autocorrelation was the greatest.	*M* = $5.68X10^{283}$; Range: [0.03, $9.34x10^{285}$]. No medium or larger effects were found.

Table 8 (Continued)

Summary of Results for Variance Components

	Bias	RMSE	Interval Coverage	Interval Width
Level-1				
Residual Variance	M = 0.0109, Range: [-0.204, 3.23]. Tended to be overestimated for all of the models except for the ID model (underestimated). As the level of the autocorrelation increased, then the bias tended to increase for all models with the exception of the ARMA(1,1) model. For the ARMA(1,1) model, the level-one variance estimate tended to be most bias when there was no correlation in level-one.	*M* = 0.22 Range [0.02, 1.19] The RMSE values decreased with an increase in the level-one sample size and total variance. The RMSE values tended to increase for the simpler models with increased autocorrelation.	*M* = 0.66 Range [0.00, 0.96]. Tended to undercover, and this was magnified to even lower coverage with increased level-three sample size and level of autocorrelation. Interval coverage was improved for longer series length for only the ID model.	*M* = $9.86\text{x}10^{282}$; Range: [0.084, $1.69\text{x}10^{284}$]. Widths decreased or became more narrow as the sample size for each level increased. For the AR(1) model, the interval widths decreased with increased level of autocorrelation.

Table 8 (Continued)
Summary of Results for Variance Components

	Bias	RMSE	Interval Coverage	Interval Width
Autocorrelation parameter	M = -0.007, Range: [-0.2591, 0.1633]. Tended to be minimal when most of the variance was at the upper levels. Tended to be underestimated by all of the models when most of the variance was at level. This was magnified for the ARMA(1,1) model.	*M* = -0.17, Range: [0.02, 0.84]. The level of autocorrelation did not impact the AR(1) model and the LRT model. However, the RMSE value tended to decreased with increased level of autocorrelation.	*M* = 0.89, Range: [0.42, 0.96]. For the AR(1) model, coverage approached nominal value; For the fit-selected models, tended to approach nominal value when there was no autocorrelation; For the ARMA (1,1) model, approached nominal value with increased level of autocorrelation.	*M* = 0.32; Range: [0.008, 2.52]. For the correctly specified AR(1) model, the interval widths were not impacted across the design factors. For the fit selected models, the widths tended to get wider for increased level of autocorrelation. For the ARMA(1,1) model, the widths were narrowest for the highest levels of autocorrelation.

CHAPTER FIVE: DISCUSSION

This chapter outlines a summary of the study and results, along with a discussion of the findings, limitations of the study, and implications for future research.

Summary of the study

Purpose

The purpose of the study was two-fold: 1) to determine the extent to which the various fit indices can correctly identify the level-one covariance structure; and 2) to investigate the effect of various forms of misspecification of the level-one error structure when using a three-level meta-analytic single-case model.

Research Questions

1. To what extent do fit indices (log likelihood ratio test, AIC, AIC corrected, BIC) correctly identify level-one covariance structure when using a three-level meta-analytic single-case model?
2. To what extent are the **fixed effect** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of treatment effect), and analysis factors (form of specification)?
3. To what extent are **confidence interval width and coverage for the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors

(number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of treatment effect), and analysis factors (form of specification)?

4. To what extent are the **Type I error and power for the test of the fixed effects** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of treatment effect), and analysis factors (form of specification)?
5. To what extent are the **variance component** parameter estimates from a three-level meta-analytic single-case model biased as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of treatment effect), and analysis factors (form of specification)?
6. To what extent are **confidence interval width and coverage for the variance components** from a three-level meta-analytic single-case model affected as a function of design factors (number of primary studies per meta-analysis, number of participants per primary study, series length per primary study), data factors (variances of the error terms, covariance structures, level of treatment effect), and analysis factors (form of specification)?

Method

Monte Carlo simulation methods were used to address the aforementioned research questions. Multiple design, data, and analysis factors were manipulated in the study. The study used a 2x2x2x2x2x5x7 factorial design. Seven experimental variables were manipulated in this study. 1) The number of primary studies per meta-analysis (10 and 30); 2) The number of participants per primary study (4 and 8); 3)The series length per participant (10 and 20); 4)Variances of the error terms (most of the variance at level-one: [σ^2=1; Σ_u = 0.5, 0.05, 0.5, 0.05; Σ_v = 0.5, 0.05, 0.5, 0.05] and most of the variance at the upper levels: [σ^2=1; Σ_u = 2, 0.2, 2, 0.2; Σ_v = 2, 0.2, 2, 0.2]); 5) Levels for the treatment effects[shift in level: 0 and 2.0; shift in slopes [0 and 0.2]; 6) the level of autocorrelation and the moving average parameter, respectively: [(0,0), (.2, 0), (.4,0), (.2, .2), (.4, .4)]; and 7) The form of model specification [i.e. ID, AR(1), ARMA (1,1)], and error structure selected by LRT, AIC, AICC, and the BIC. For each of the 96 data and design conditions, 5000 simulated data sets were generated using SAS IML (SAS Institute, Inc., 2008). These data sets were then specified using the a priori model selection of the level-one error structure and the use of fit criteria (post hoc model selection) of the level-one error structure.

This study first examined the proportion of times that each fit index correctly selected the appropriate model. Secondly, this study examined the treatment effects (i.e., the overall average treatment effect and the overall average difference between baseline and treatment slope) and the variance components (e.g., the between-person within-study variance in the average treatment effect, the between-person within-study variance in the average difference between baseline and treatment slope, the between-study variance in the overall average treatment effect, and the between-study variance in the average difference between the baseline and the treatment slopes) in a multi-level model.

Discussion of Study Results

Correct Model Selection

Results indicated that the proportion of times that the ID model was correctly identified was greatest for the models selected by the LRT and least for the models selected by the AIC. The variability was then explored by running GLM models to identify medium or larger effects. The model revealed that the interaction effect of the number of primary studies included in the meta-analysis and the type of fit index had an impact on the proportion of times that the ID model was correctly selected. The relationship revealed that the proportion of times that the ID model was correctly specified increased when the number of primary studies included in the meta-analysis increased. However, this increase was not identical across all models; specifically, the improvement was greatest for the BIC fit index and least for the models selected by the LRT. Past research suggested that the BIC had better performance with increased sample size (Raftery, 1995), and given that that the SAS PROC MIXED uses the number of independent sampling units as the sample size, and in this case, this would imply the number of studies used in the meta-analysis. That could be why the noted increased performance with the BIC fit index when the number of studies in the meta-analysis (independent sampling units) increased.

The proportion of times that the AR (1) model was correctly identified was then explored and revealed that on average the AR (1) model was correctly selected most often by the LRT, and the least often by the AIC fit index. The results of the GLM models then found that there were three medium or larger effects: the series length, the number of studies to be included in the meta-analysis, and the type of fit index that was used for selection. When using the three-level model, all of the fit indices correctly identified the AR (1) model at least 80% of the time. Specifically, the relationship revealed that as the series length increased from 10 to 20, then the proportion of times that the AR (1) model was correctly identified also increased. Similarly, as the number of primary studies to be included in the meta-analysis increased, then the proportion of times that the AR (1) model was selected correctly also increased. Lastly, the association of the proportion of times that the AR(1) was correctly selected greatly depended on the fit index. The LRT extremely outperformed the other fit indices, followed by the BIC, and finally the AICC and the AIC. Previous work with the two-level models (Ferron, Dailey, & Yi, 2002; Kesselman, Algina, Kowalchuk, & Wolfinger, 1999) had suggested that overall the fit indices did not perform well in terms of model selection; explicitly, Ferron, Dailey, and Yi (2002) found that the AIC only correctly identified the models 47% of the time. Additionally, they found that upper level sample size mattered more when there were shorter series.

The distribution for the proportion of times that the ARMA (1,1) model was correctly identified was examined. The results revealed that none of the fit indices correctly selected the model more than 20% of the times. The results also indicated that the greatest mean proportion of times was for the AIC (M = 0.19), meanwhile the LRT (M = 0.07) had the smallest proportion of correct identification for this model. One possible explanation for the low identification rates for the LRT is that the log likelihood ratio test would have to reject multiple significant tests in order to correctly identify the ARMA(1,1) model. The association of the mean proportion of times that the ARMA(1,1) model was correctly identified by the fit indices and the interaction effect of the type of fit index and the number of primary studies to be included in the meta-analysis was then explored. Concretely, for the BIC fit index, the greatest decrease in the correct identification was seen when the number of primary studies decreased. This seemed counterintuitive, given the expected improvement with the upper-level sample size increase. This can be due to the fact that, overall (less than 20% of the time), the fit indices did not correctly select the ARMA (1,1) model.

However, there was hardly any difference for the models selected by the LRT in the mean proportion of correct selection when the number of primary studies increased. Additionally, Gomez, Schaalje, and Fellingham (2005) found that success rates tend to rely greatly on sample size and type of covariance structure; rates tend to be higher for the simpler covariance structures. Similar results were found in this current study in which the success rates tended to be higher for the models which had less complex error structures for particular fit indices.

Fixed Effects

The fixed effects were examined in terms of various outcomes of interest: bias, RMSE, confidence interval coverage and width, Type I error, and power for the tests of the fixed effects. The extent to which the fixed effects were biased as a function of the study's design factors was examined by looking at two outcomes of interest, the bias and the RMSE. The results indicated that for both of the treatment effects, the shift in level and the shift in slopes, the average bias value was close to zero, across all of the combinations of the design factors.

An examination of the RMSE values revealed similar results for both of the treatment effects across most of the design factors. However, the RMSE values were impacted by the number of primary studies included in the meta-analysis (the RMSE values decreased as the number of primary studies increased) and by the variances of the error terms (as the variance shifted from most of the variance at the upper levels to most of the variance at level one), the RMSE values tended to decrease. This outcome suggest that if possible, researchers should strive to increase their level-three sample size (number of primary studies included in the meta-analysis).

An initial analysis of the confidence interval coverage revealed that there was no meaningful variability in the mean interval coverage for the fixed effects. Therefore, no further investigation was warranted. Prior research (Ferron et al., 2009) had shown that the coverage estimates tended to be highest, 0.942 when autocorrelation was modeled versus when the autocorrelation was not modeled. However, the current study illustrated that the mean coverage approached the nominal value across all seven models for both of the fixed effects (shift in level and shift in slope).

An exploration of the confidence interval widths indicated that as the number of primary studies increased and the variances of the error terms shifted to being mostly at level one instead of at the upper levels, then the mean interval width tended to decrease. Additionally, for the ARMA (1,1) model, the mean interval width vastly decreased as the level of the autocorrelation parameter increased. These results indicated again for applied researchers to attempt to add to the number of primary studies included in the meta-analysis when possible. This was consistent with previous work that investigated the three-level model (Owens, 2011), which found that the interval widths tended to be smallest when the number of primary studies was largest and when most of the variance was at level one as opposed to most of the variance being at the upper levels. The results also supported the findings of previous work which looked at the two-level model (Ferron et al., 2009), which showed that the mean interval widths tended to be smallest when there were more upper level units (number of participants) and there was less variability among the upper level units.

The exploration of the Type I error rates indicated that the Type I error fell within Bradley's (1978) liberal criterion for both of the fixed effects. Therefore, no further analyses were warranted to examine the variability of the mean Type I error rates. These results are slightly different from prior studies (Gomez, Schaalje, & Fellingham, 2005) which found that the Type I error rates tended to be higher for the models selected solely by the AIC and BIC. However, this current study found that the Type I error tended to be close to the nominal value of 0.05 across all seven models.

Power estimates for the phase effect (shift in level) and the interaction effect (shift in slopes) revealed that when most of the variance is at level one, then the power estimates are greater than 0.9 and did not tend to depend on the number of primary studies included in the meta-analysis. However, when most of the variance is at the upper levels, and the number of primary studies is increased, then the mean power estimate also increased. Previous research did not look directly at power estimates; however the conclusions regarding the importance of increasing the upper level units can still be noted. This study showed that this becomes increasingly important when there is great variability at the upper levels.

An analysis of under-, over-, and correct specification was done to investigate whether there was a general rule that can used when selecting a level-one error structure. Specifically, for the bias and 95% confidence interval coverage was examined. The analysis found that it really did not make a huge difference whether the level-one structure was correctly specified, over-specified, or underspecified. Therefore, no general rule of thumb could be applied in terms of the fixed effects.

Variance components

Variance components were then analyzed in terms of bias, RMSE, confidence interval coverage and widths. First, the bias for both level-three and level-two variance components were examined for the treatment effects. Relative bias was also calculated for these parameters since their known values were not equal to 1 and the parameter did not contain levels that included the value of 0. An exploration of the level-three and level-two variance components revealed that although, the estimates tended to be overestimated, there were no medium or larger effects. Previous work had revealed that there was substantial bias in the variance components when the number of participants were small and the series length was short, either 4 or 8 (Kwok, West, & Green, 2007; Murphy & Pituch, 2009) even when the model was correctly specified. Additionally, Owens (2011) concluded that the when most of the variance was at the upper levels, then there was increasingly more bias in the variance components. However, this current study did not show any of the design factors having a medium or larger effect on the bias for the level two and three variance components with the original data. Due to the large range for the variance, the data were then trimmed for the relative bias and the RMSE values; the results of the additional analyses are contained in Appendix A.

However, the residual variance tended to be overestimated for the majority of the models (all of the models with the exception of the ID model). For the ID model, the bias in the residual variance tended to be underestimated. As the level of the autocorrelation parameter increased, then the bias in the level-one variance became increasingly larger for all of the models with the exception of the ARMA (1,1) model. The residual variance estimate for the ARMA (1,1) model tended to be most bias when there was no autocorrelation and least bias when the autocorrelation parameter was 0.2.

The RMSE for the level-one variance revealed similar results. The RMSE values tended to increase with greater levels of autocorrelation and decrease with an increase in the series length and total variance for all of the models, except again, for the ARMA (1, 1) model. Some of these conclusions support previous work with the three-level model (Owens, 2011) which indicated that the bias in the residual variance was dependent on the autocorrelation parameter. This would seem intuitive given that the autocorrelation parameter represents the correlation, or the relationship between the observations within a participant. The more correlated these errors are, then it could be expected that there would be more difficulty in producing precise parameter estimates for the level-one variance. Previous work investigating the three-level model did not look at the ARMA(1,1) models. The current study found that when there is no autocorrelation, then the ARMA(1,1) model was problematic in estimating the level-one variance. The model tends to be most precise when there is at least a moderate amount of correlation among the level-one errors (both ρ and □ at least 0.2). This also seems instinctive, given that the ARMA(1,1) model is attempting to estimate a more complex correlated level-one error structure. When there is no correlation among the level-one error, then the model's parameter estimates tend to be problematic.

The bias and RMSE found in the autocorrelation parameter was then explored across all of the design factors. The interaction between the variances of the error terms and the type of model, tended to affect the bias observed for the autocorrelation parameter. When most of the variance was at the upper levels, the mean bias for the autocorrelation parameter tended to be minimal across the models. Similarly, when most of the variance was at level one, then the autocorrelation parameter estimate tended to be slightly underestimated for all of the models, and again, this was magnified for the ARMA(1,1) model, when there was no autocorrelation. The study did not completely support other work investigating the three-level model (Owens, 2011), which found that on average, the autocorrelation parameter tended to be unbiased across all factors. Conversely, this study did indicate that the variances of the error terms did seem to impact the precision of the autocorrelation parameter.

This current study indicated that the moving average parameter tended to be underestimated by all of the fit-selected models, but overestimated for the ARMA (1,1) model. The bias found in the moving average parameter tended to be greatly impacted by the amount of correlation found in the level-one error structure. The parameter estimate was overestimated when there was no correlation, slightly underestimated when the moving average parameter was 0.2, and even more underestimated as the moving average parameter increased to 0.4. Again, this result indicates that there should be at least some moderate level of correlation among the error structure in order to observe optimal performance when utilizing a model as complex as the ARMA (1,1) model.

Confidence interval coverage for each of the variance components were estimated as the proportion of the confidence intervals at the .95 level that contained the true parameter estimates. Coverage intervals for the level-three variance component for the phase effect tended to overcover but were closest to the nominal value of 0.95 when the number of primary studies was increased from 10 to 30. Further examination of the effects revealed that the relationship between interval coverage and the variances of the error terms depended on the number of participants. Furthermore, when most of the variance was at the upper levels, the impact of the level-two sample size was attenuated. However, when most of the variance of the error terms is at level one, the mean interval coverage greatly increased as the number of participants increased. Specifically, when most of the variance is at the upper levels, regardless of the level-two sample size, then the mean interval coverage would tend to overcover, or would be higher than the nominal value of 0.95. When most of the variance is at level one, and the number of participants is 4, then the interval coverage tend to undercover. If the number of participants is increased to 8, then the interval coverage tend to overcover.

Interval coverage for the interaction effect for the level-three variance component was then investigated. When most of the variance of the error terms is at the upper levels, and the series length was 10, then the interval tend to slightly overcover, and this was magnified when the series length was increased to 20. However, when most of the variance is at level one, and the series length was 10, then the interval would undercover, and when the series length was increased to 20, then the coverage would tend to overcover. There was little to no difference between the levels of the error variances when the series length was 20. Additionally, when the number of participants was 4, the mean interval coverage for the interaction effect tended to undercover, and the inverse was observed, the mean interval coverage tended to overcover when the number of participants was 8. Prior research involving the three-level meta-analytic model (Owens, 2011) also indicated that the coverage for the level-three variance components tended to be greater than the nominal value of 0.95. The research also showed that similar impact factors: the combinations of sample sizes at each of the levels and the variances of the error terms.

Interval coverage for the level-two variance components for both of the intervention effects tended to slightly undercover. A further examination into the medium or larger effects for the level-three variance component for the phase effect (shift in level) indicated that when the most of the variance of the error terms is at level one, the mean interval coverage tended to increasingly undercover, or fall below the nominal value of 0.95. This undercoverage was magnified for the ID model, with interval coverage approximately 0.60. Similarly, the analysis revealed that for all of the models, with the exception of the ID model, the level of autocorrelation did not affect coverage. However, for the ID model, the mean interval coverage was lowest when the autocorrelation parameter was greatest. There were instances when the coverage was close to 0 for the ID model. Additionally, for the ID model, coverage approached the nominal value of 0.95 when there was no autocorrelation. Similar results were found for the level-two variance component for the shift in slopes (interaction effect). This was similar to previous work analyzing the three-level models (Owens, 2011) which also found that the level-two variance components tended to undercover.

Confidence interval coverage for level-one residual variance was then analyzed as a function of the design factors in the study. The coverage was problematic, ranging from a mean of 0.56 to 0.70 across the various models. Only the medium or larger effects were examined and found that the mean interval coverage for the residual variance decreased to even lower values as both the level of autocorrelation and the number of primary studies to be included in the meta-analysis increased. Additionally, the interval coverage for the level-one variance was lower when the series length was longer for all models with the exception of the ID model. The reverse was true for the ID model, which showed an increase in the mean interval width when the series length was longer. Prior work involving the three-level model (Owens, 2011) also found that the interval coverage for the level-one variance was lower than the nominal rate of 0.95. An additional finding in prior works (Owens, 2011) showed that when the level of autocorrelation was zero, then the coverage rates were optimal, which is consistent with this current study. However, this study also found that again, the combination of the number of primary studies and series length also had an impact on the interval coverage rates for the residual variance.

The distribution for the coverage for the autocorrelation parameter illustrated great variability both across and within the models. A further examination into the medium or larger effects revealed that for the AR(1) model, the mean interval coverage was similar (approached the nominal value of 0.95) across all levels of the autocorrelation parameter. The fit selected models revealed that the mean interval coverage was lower than the nominal value with moderate levels of autocorrelation ($\rho = 0.2$) and approached the nominal value when there was no autocorrelation ($\rho = 0.0$). Finally, the ARMA(1,1) model illustrated the inverse relationship, the mean interval coverage increased as the level of the autocorrelation parameter increased from 0.0 to 0.4. When the level of the autocorrelation parameter was greatest, then the interval coverage for the ARMA (1,1) model approached nominal value. These findings again supports prior work (Gomez, Schaalje, & Fellingham, 2005) and revealed the fit selected models tend to perform better with the less complex error structures, while the ARMA(1,1) tend to favor the more complicated error structures.

Mean interval coverage was then analyzed for the moving average parameter. Additionally, the medium or larger effects illustrated that the mean interval coverage decreased as the level of the autocorrelation parameter increased from 0.0 to 0.2, but remained comparable from $\rho = 0.2$ to $\rho = 0.4$. However, a more in-depth analysis revealed the effect of the level of the moving average parameter on the mean interval coverage depended on the type of model. Specifically, the interval coverage was greatest for the fit index selected models when the level of the moving average parameter was 0.0, and smallest when the moving average parameter was 0.2 and 0.4. However, for the ARMA (1,1) model, the mean interval coverage was lowest when the moving average parameter was 0.0 and greatest when the parameter was 0.2 and 0.4. This conclusion reinforces a common trend in this chapter: the fit index selected models tend to have optimal performance for less complex models, while the ARMA(1,1) performs best for more complicated error structures.

Confidence interval width was described as the average difference between the upper and the lower limits of the 95% confidence intervals. The confidence interval width for the both the level-three and level-two variance components were analyzed. The widths were large, this finding is consistent with prior work dealing with both the two-level (Ferron et al., 2009) and the three-level (Owens, 2011) models. Interval widths for the level-one variance revealed that the mean interval width decreased as the series length, number of participants, and number of studies to be included in meta-analysis increased. Previous work (Owens, 2011) found similar results that the widths became even smaller with increased sample size at each level. Additionally, Owens (2011) had found that the level of the autocorrelation parameter also affected the interval width. This current study did find a similar conclusion, however also finding that this depended on the type of model. More specifically, for the AR (1) model, the width tended to decrease as the level of the autocorrelation increased. There was no impact on the width for the ID model with varying levels of autocorrelation.

Confidence interval width for the autocorrelation parameter was impacted by the interaction effect: the type of model that was used to estimate the parameters and the level of the autocorrelation parameter. For the correctly specified first-order autoregressive model, AR (1), there was little difference in the mean interval width. The models selected by the fit indices revealed that the width were narrowest when there was no autocorrelation and wider for the higher levels of autocorrelation. The ARMA (1,1) model illustrated, again, the better accuracy for the higher levels of autocorrelation. The mean interval width was greatest (M = 8.30) for the correctly specified ARMA (1, 1) model, which seemed again kind of counter intuitive. One possible explanation again, could be the fact that the fit selected models rarely correctly identified (less than 20% of the time) the ARMA (1,1) model, therefore rarely estimating the moving average parameter. This can be coupled with the fact that the ARMA (1, 1) model is not precise, however, it is estimating the model more often than the fit-selected models. The interval width decreased greatly for the ARMA (1, 1) model as the level of moving average parameter increased. This again, maintained the notion that the performance of the ARMA (1, 1) model is greatly improved with the presence of a more correlated level-one error structure.

The analysis for the both the level-three and level-two variance components revealed that the bias was comparable across the models for these variance components. However, there seemed to be a difference in the estimation of the residual variance based on whether the level-one error structure was correctly, under-, or over- specified. Specifically, the analysis revealed that for the residual variance, when the ID model is the correct model, the bias was minimal for the ID model, but greatest for the ARMA (1,1) model. When AR(1) was the correct model for the residual variance, the bias was comparable for the under-specified (ID model), and the correctly-specified AR(1) model. However, the ARMA(1,1) model did not do a comparable job estimating the residual variance. This conclusion again supporting the finding that the ARMA (1,1) model tended to perform worse with little to no autocorrelation.

Limitations of the Study

There are many benefits to conducting Monte Carlo or simulation research. These types of studies allow researchers to operate under the true parameter values and determine how various design factors or values for these factors can impact the true parameter estimates. The conditions, that is, the design factors and the values chosen for each of those factors, affect the study's generalizability.

The data in this study were simulated based on specific design conditions. Those conditions were chosen based on a review of single-case literature, meta-analyses of single-case data, and applied work that was done using the three-level model to aggregate data across studies. The specific conditions chosen for this study are only a portion of the possible options that could have been included in this current study. Therefore, the results of this study can only be generalized to studies with the same or similar conditions. Any conclusions beyond the observed conditions should be interpreted with caution. The next section will address detailed limitations based on the specific design factors that were used in this study.

First, the study assumed that all of the primary studies included in the meta-analysis used a multiple-baseline design. This design was selected over the previously discussed reversal (or A-B-A-B) design that was used in the social behavior study (Lorimer & Simpson, 2002) or even the popular alternating treatment design (Kazdin, 2009; Shadish & Sullivan, 2011). An additional feature of the single case studies was that the dependent variable was assumed to be continuous for all of the studies. The use of continuous variables in single-case studies is common in terms of mathematics achievement (Billingsley, Scheuermann, & Webber, 2009) or words read per minute (Tam, Heward, & Heng, 2006). There are various types of outcomes that are commonly used in single case studies, such as binary, ordinal, or count outcomes, for example, counting the number of times that a student talks out without raising their hands or the number of times that a student leaves their seat. These examples would require different types of assumptions using a Poisson distribution (Shadish & Rindskopf, 2007; Shadish et al., 2008).

This study assumed that the same outcome was used across studies. This is a huge assumption considering that outcomes can be measured in a variety of ways. For example, there are many measures that can be used to appropriately measure mathematics achievement.

Additionally, the models that were used to analyze the data only included linear trends, however, more complex trends, such as adding a quadratic or cubic term, could have been used; non-linear trends are also commonly used to investigate single case data (Beretvas, Hembry, Van den Noortgate, & Ferron, 2013; Shadish & Rindskopf, 2007). In addition to assuming the same outcome, the study used the raw data in the synthesis of the study, perhaps there could be different results if the data were standardized instead.

This study also found that overall, the treatment effects, including both the shift in level and the shift in slopes, were not biased. However, this study did not look at the effects on particular groups of individuals, such as boys vs. girls. This would involve conducting some moderator analyses.

Implications for Researchers, Meta-analysts, and Methodologists

Meta-analysis of single-case studies has become increasingly popular, due to many elements. Accountability and the need to associate a study with an effect have led to statistical methods being applied, in addition to the popular visual analysis of single case data. The study involves not only look at an intervention within a study, but examines a method for combining treatment effects across multiple studies using the raw data from single-case studies. This study has various implications, not only for the applied researchers who are conducting intervention research daily; but also for the meta-analysts, who seek to investigate intervention effectiveness across multiple studies. Additionally, this study has significance for the methodologists who seek precise methods for determining treatment effects when meta-analyzing single-case research.

The results of this study can also be applied beyond the framework for this model. The results can be generalized to many three-level models. This can include, but is not limited to most longitudinal studies, that may have multiple observations nested within an individual, individuals nested within studies or schools or even, classrooms.

Implications for the Applied Single-Case Researcher

First, this study examined the use of fit indices to correctly identify covariance structures and found that certain fit indices performed optimal under a range of conditions. For the ID model (assuming the data has no autocorrelation), the LRT tended to have the best performance. The AR(1) model was correctly identified most often by the LRT index, however performance across all of the models was improved with increased series length and increasing the number of primary studies. All of the fit indices on average, correctly identified the AR(1) model at least 80% of the times.

The overall performance of the fit indices for selecting the ARMA(1,1) model was not as positive. None of the fit indices used in this current study correctly identified the ARMA(1,1) model more than 20% of the time. However, if one of the aforementioned indices must be used to identify the ARMA(1,1) model, then choose the AIC, for the AIC correctly selected the model the most often and the LRT fit index selected the model the least often. These findings indicate that fit indices work well for the less complicated error structure (ID and AR(1)), however if the researcher believe the data has a more complex error structure, such as ARMA(1,1) then fit indices is not be a suitable option for correct model selection. The researcher should just select the level-one error structure a priori. The conclusions also suggest that increasing the series length improved performance of the fit indices across all models. Given this conclusion, it is recommended that if researchers would like to use fit indices for model selection, then increasing the number of observations would increase the precision of correct model identification.

Another factor that continued to impact parameter estimates was the amount of total variance in the study. This impact can be reduced if applied researchers can attempt to control the amount of extraneous variability in the study that may be due to the study's design. One example would be to ensure that there is stable performance in the baseline phase. Kazdin (2011) characterizes stable performance as one that is absent of trend or slope with little to no variability. Additionally, results revealed that increasing the number of participants in the study could greatly reduce the bias observed in the variance components.

Implications for the Applied Single-Case Meta-analyst

Meta-analysts who are interested in the treatment effects (both the shift in levels and the shift in slopes), the results are promising. Overall, the treatment effects were not biased across the studies, however more precise estimates can be obtained with increasing number of studies to be included in the meta-analysis. The treatment effects are sometimes, of most value to answer the question as to whether or not the intervention was effective across studies. However, maybe work can be done in terms of looking at potentially adding moderators to gain a deeper understanding of the treatment effects across particular subgroups or factors. Additionally, the level of autocorrelation only affected the ARMA (1,1) model: more precise estimates were found with increasing levels of autocorrelation for the ARMA(1,1) model. This seems again, intuitive given that this model is trying to estimate a very complex correlated error structure, and if the error structure is not correlated, then the ARMA (1,1) model should not be used. The current study supported past research and revealed that the variance components (both level-three and level-two) were biased across all design factors. However, slight improvements were seen when the number of studies to be included in the meta-analysis were increased. This would imply that when possible, meta-analyst should attempt to increase the number of studies to be included in the meta-analysis.

The level-one variance components (residual variance, autocorrelation parameter, and the moving average parameter) were biased, and the precision of the estimates heavily depended on the level of autocorrelation that was used in the model. The residual variance became more bias with increased autocorrelation; conversely, the moving average parameter became less biased with increased autocorrelation. The precision of the autocorrelation parameter was unaffected by varying levels of the correlated errors. This suggests to meta-analysts that the presence of a correlated vs. uncorrelated level-one error structure should determine which model one chooses for their data. Consequently, this can greatly impact the accuracy of the parameter estimates and thus, inferences that can be made from these estimates. Therefore, if the meta-analyst hypothesizes that the error structure is uncorrelated (similar to an ID model) or has a simple correlated error structure, then the fit indices are reliable for selection of the correct error structure. However, if the error structure is assumed to be more complex, then selecting this error structure a priori is better than relying on the fit indices.

Implications for Methodologists

The study analyzed the use of a range of violations to the independence error assumption. Additionally, the study examined the accuracy of parameter estimates when the models were misspecified. Overall, the model was robust to many of the misspecifications in terms of the treatment effects. However, for the variance components, the model did not perform as well. It appeared that the model would randomly malfunction, and occasionally estimate the variance to be large, which led to the average parameter estimates being biased. Further work should be done using different estimation methods, such as the Bayesian approach (Baldwin & Fellingham, 2013; Gelman, 2006) to see if the observed bias in the variance components can be reduced.

Additionally, methodologists may want to look at coupling the violation of the error assumption with other violations such as also investigating non-normal distributions. This would be a reasonable study given that many of the data in single case research is not continuous and normally distributed. Lastly, more simulation work can be done to see if the performance of the fit indices would improve is there was cutoff instead of just looking at relative differences when utilizing the fit indices to correctly identify the level-one error structure (e.g. AIC smaller by at least 2 or 3).

REFERENCES

Allison, D.B. & Gorman, B.S. (1993). Calculating effect sizes for meta-analysis: The case of the single-case. *Behavior, Research, and Therapy, 31*, 621-631.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on automatic control, AC-19,* 716-723.

Baek, E., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45, 65-74.

Baek, E.K., Moeyaert,M., Petit-Bois, M., Beretvas, S.N., Van den Noortgate, W., & Ferron, J.M., (2013). The use of multilevel analysis for intergrating single-case experimental design results within a study and across studies. *Neuropsychological Rehabilitation: An International Journal*, doi: 10.1080/09602011.2013.835740.

Baek, E.K., Petit-Bois, M., Ferron, J.M. (2012, April). *The effect of error structure specification on the meta-analysis of single-case studies of reading fluency interventions.* Poster presented at the Annual Meeting of the American Educational Research Association. Vancouver, CA.

Barlow, D., Nock, M., Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Pearson Education, Inc.

Baldwin, S.A. & Fellingham, G.W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18 (2),* 151-164.

Beretvas, S., & Chung, H. A Review of meta-analyses of single subject experimental designs: Methodological issues and practice. *Informa Healthcare USA, Inc., 129-141*. doi: 10.1080./17489530802446302

Brossart, D.F., Parker, R.I., Olson, E.A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*(5), 531-563. doi: 10.1177/0145445503261167

Bellini, S., Peters, J.K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. *Remedial and Special Education, 28*(3), 153-162.

Codding, R.S., Burns, M.K., & Lukito, G. (2011) Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*(1), 36-47.

Busk, P.L., & Marascuilo, L.A. (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment, 10*, 229-242.

Center, B.A., Skiba, R.J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intrasubject design research. *Journal of Special Education, 19*, 387-400.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum Associates.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Longitudinal data analysis*. New York: Oxford University Press Inc.

Edgington, E.S. (2010). Randomization tests for one-subject operant experiments. *The Journal of Pscyhology: Interdisciplinary and Applied, 90 (1),* 57-68.

Faith, M.S., Allison, D.B., & Gorman, B.S. (1996). Meta-analysis of single-case research. In R.D. Franklin, D.B. Allison, B.S. Gorman, *Design and analysis of single-case research* (pp. 245-277). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Farmer, J.L., Owens, C.M., Ferron, J.M., & Allsopp, D.H. (2010, April). *A methodological review of single-case meta-analyses.* Paper presented at the American Educational Research Association. Denver, CO.

Ferron, J.M., Bell, B.A., Hess, M.R., Rendina-Giobioff, G. & Hibbard, S.T. (2009). Making treatment effect inferences from multiple baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods, 41*(2), 372-384.

Ferron, J., Dailey, R., & Yi. Q. (2002). Effects of misspecifying the first level error structure in two-level models of change. *Multivariate Behavioral Research, 37*(3), 379-403.

Ferron, J.M, Farmer, J.L., & Owens, C.M. (2010). Estimating individual treatment effects from multiple-baseline data: A monte carlo study of multilevel-modeling approaches. *Behavior Research Methods, 42* (4), 930-943.

Ferron, J., & Onghena P. (1996). The power of randomization tests for single-case phase designs. *The Journal of Experimental Education*, *64*(3), 21-239.

Fisch, G.S. (2001). Evaluating data from behavioral analysis: visual inspection or statistical models? *Behavior Processes*, *54*, 137-154.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1(3)*, 515-533.

Gentile, J.R., Roden, A.H., & Klein, R.D. (1972). An analysis of variance model for intra-subject replication designs. *Journal of Applied Behavior Analysis, 5,* 193-198.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5,* 3-8.

Gorsuch, R.L. (1983). Three methods for anlayzing limited time-series (N of 1) data. *Behavioral Assessment, 5*, 141-154.

Goldstein, H., &Yang, M. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society, Series C: Applied Statistics, 49*, 339-412.

Gomez, E. V. (2005). *Performance of the kenward-roger method when the covariance strucure is selected using AIC and BIC*. (Unpublished doctoral dissertation). Brigham Young University: Provo, Utah.

Gorman, B.S., & Allison, D.B. (1996). Statistical Alternatives for Single-Case Designs. In R.D. Franklin, D.B. Allison, B.S. Gorman, *Design and Analysis of Single-Case Research* (pp. 159-214). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hedges, L., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Hoogland, J.J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research, 26* , 328-367.

Huitema, B.E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment, 7,* 107-118.

Huitema, B.E., & McKean, J.W. (1991). Autocorrelation estimation and inference with small samples. *Pscyhological Bulletin, 110*(2), 291-304.

Huitema, B.E., & McKean, J.W. (1998). Irrelevant Autocorrelation in Least-Squares Intervention Models. *Psychological Methods, 3*(1), 104-116.

Huitema, B.E., & McKean, J.W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60* (1), 38-58. doi: 10.1177/00131640021970358

Hox, J. J., & de Leeuw, E. D. (1997). Meta-analyse via multiniveau-modellen. *Tijdschrift Voor Onderwijsresearch*, *22*, 16-28.

Jones, R.R., Weinrott, M.R., & Vaught, R.S. (1978). Effects of serial D=dependency on the agreement between visual analysis and statistical inference. *Journal of Applied Behavior Analysis*, *11* (2), 277-283.

Kazdin, A.E. (2011). *Single-case research designs: Methods for clinical and applied settings (*2nd ed.). New York, New York: Oxford University Press.

Kesselman, H., Algina, J. Kowalchuk, R. & Wolfinger, R. (1999). A comparison of two approaches for selecting covariance structures in the Analysis of repeated measurements. SAS Institute.

Kratochwill, T.R. & Levin, J.R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15* (2), 124-144.

Kratochwill, T., Alden, K., Demuth, D., Dawson, D., Panicucci, C., Arntson, P., McMurray, N., Hempstead, J., & Levin, J. (1974). A further consideration in the application of an analysis of variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis, 7*(4), 629-633.

Kromrey, J.D., & Foster-Johnson, L.(1996). Determining the efficacy of intervention: The use of effect sizes for data analysis in single-subject research. *Journal of Experimental Education, 65* (1), 73-93.

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

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

Liu, S., Rovine, M. & Molenaar, P. (2012). Using fit indexes to select a covariance model for longitudinal data. *Structural Equation Modeling*, *19*, 633-650. doi:10.1080/10705511.2012.726918

Lorimer, P.A., & Simpson, R.L. (2002). The use of social stories as a preventable behavior intervention in a home setting with a child with autism. *Journal of Positive Behavior Interventions, 4*(1), 53-60.

Ma, H.H. (2006). An alternative method for quantitative synthesis of single-subject research: Percentage of data points exceeding the median. *Behavior Modification*, *30,* 598-617.

Maggin, D.M, Swaminathan, H., Rogers, H., O'Keeffee, B. Sugai, G., & Horner, R. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, *49*, 301-321.

Matyas, T.A. & Greenwood, K.M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis*, *23* (3), 341-351.

Matyas, T.A., & Greenwood, K.M. (1991). Problems in the estimation of autocorrelation and some implications for behavioral data. *Behavioral Assessment*, *13*, 137-157.

Matyas, T.A., & Greenwood, K.M. (1996). Serial dependency in single-case time series. In R.D. Franklin, D.B. Allison, & B.S. Gorman, *Design and analysis of single-case research* (pp. 215-243). Mahwah, New Jersey: Lawrence Erlbaum Associates.

McCoach, D.B. & Black, A.C. (2008). Evaluation of model fit and adequacy. In A. O'Connell & D. McCoach, Multilevel modeling of educational data (245-272). Charlotte, NC: Information Age Publishing, Inc.

Mooney, P., Ryan, J.B., Uhing, B.M., Reid, R., & Epstein, M.H. (2005). A review of self-management interventions targeting academic outcomes for students with emotional and behavioral disorders. *Journal of Behavioral Education, 14*, 303-331.

Murphy, D.L., & Pituch, K.A. (2009). The Performance of Multi-level Growth Curve Models Under an Auto regressive Moving Average Process. *The Journal of Experimental Education, 77(3)*, 255-282.

Onghena, P., & Edgington, , E.S. (2005). Customization of pain treatments: Single-case design and analysis. *The Clinical Journal of Pain, 21*, 56-68.

Ostrom, C.W., Jr. (1990). *Time series analysis: Regression techniques* (2nd ed.). Beverly Hills, CA: Sage.

Owens, C. M. (2011). Meta-analysis of single-case data: A monte carlo investigation of a three level model. (Unpublished doctoral dissertation). University of South Florida: Tampa, FL.

Owens, C.M., & Ferron, J.M. (2011). Synthesizing single-case studies: A Monte carlo examination of a three-level meta-analytic model. Behavior Research, doi: 10.3758/s13428-011-0180-y

Parker, R.I., & Brossart, D.F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy, 34*, 189-211.

Parker, R.I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, *40*, 194-204.

Parker, R., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357-367.

Petit-Bois, M. (2012, April). *Consequences of misspecification of growth trajectories when meta-analyzing single-case data using a three level model*. Poster presented at the American Educational Research Association. VanCouver, CA.

Petit-Bois, M., Baek, E.K., Ferron, J.M. (2013, April). *The consequences of specifying level-1 error structure on the multi-level meta-analysis of single-case studies: A Monte Carlo Study*. Poster presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Scruggs, T.E., Mastropieri, M.A., & Castro, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8* (2), 24-33.

Shadish, W.R., & Rindskopf, D.M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation, 113, 95-109*.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the metaanalysis of single-case experimental designs. Evidence-Based Communication Assessment and Intervention, 2, 188–196.

Shadish, W.R., & Sullivan, K.J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research, 43*, 971-980. doi:10.3578/s13428-011-0111-y

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

Sivo, S., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series processes on latent growth curve model estimates. *Structural Equation Modeling, 12*(2), 215-231.

Sivo, S.A., & Willson, V.L. (2010). Is parsimony always desirable? Identifying the correct model for a longitudinal panel data set. *The Journal of Experimental Education*, *66* (3), 249-255.

Swanson, H.L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject design intervention research for students with LD. *Journal of Learning Disabilities, 33*, 114-336.

Tabachnick, B.G. & Fidell, L.S. (2008). *Using Multivariate Statistics* (5th ed.). Pearson: Boston.

Templeton, T., Neel, R., & Blood, E. (2008). Meta-analysis of math interventions for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 16 (4), 226-239*.

Toothaker, L., Banz, M., Noble, C., Camp, J. & Davis, D. (1983). The failure of ANOVA-based tests. *Journal of Educational Statistics*, *8*(4), 289-309.

Ugille, M., Moeyaert, M., Beretvas, T., Ferron, J., & Van den Noortgate, W. (2012). Multilevel meta-analysis of single-subject experimental designs: A simulation study. *Behavior Res.*doi: 10.3758/s13428-012-0213-1

Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63* (5), 765-790.

Van Den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today, 8*(2), 196-209.

Van Den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Informa Healthcare USA, Inc., 142-151*. doi: 10.1080./17489530802446302

Wolfinger, R.D. (1996). Heterogenous variance: Covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1*(2), 205-230.

Xin, Y.P., & Jitendra, A.K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education*, *32*(4), 207-225.

APPENDIX A: TRIMMED DATA ANALYSIS

Trimmed Data Analyses for the Variance Components

Relative Bias

The next section describes the additional analyses that were conducted for the trimmed relative bias values for the variance components for each of the treatment effects across both of the upper levels for the model.

Level-three variance for the overall average treatment effect for the phase effect (shift in level). GLM models were run to determine if any of the design factors had a medium or larger effect; the effects (the number of primary studies to be included in the meta-analysis [η^2= 0.093] and the variances of the error terms [η^2= 0.096]) were identified. The model including 4-way interactions explained 95% of the variability. The figure (see Figure A1) below displays the effects and the means for the relative bias across each value for the effect. Specifically, the relative bias decreased as the number of primary studies increased from 10 (*M* = 0.01, *SD* = 0.02) to 30 (*M* = 0.002, *SD* = 0.009). The variability also tended to decrease with increased number of primary studies included in the meta-analysis.

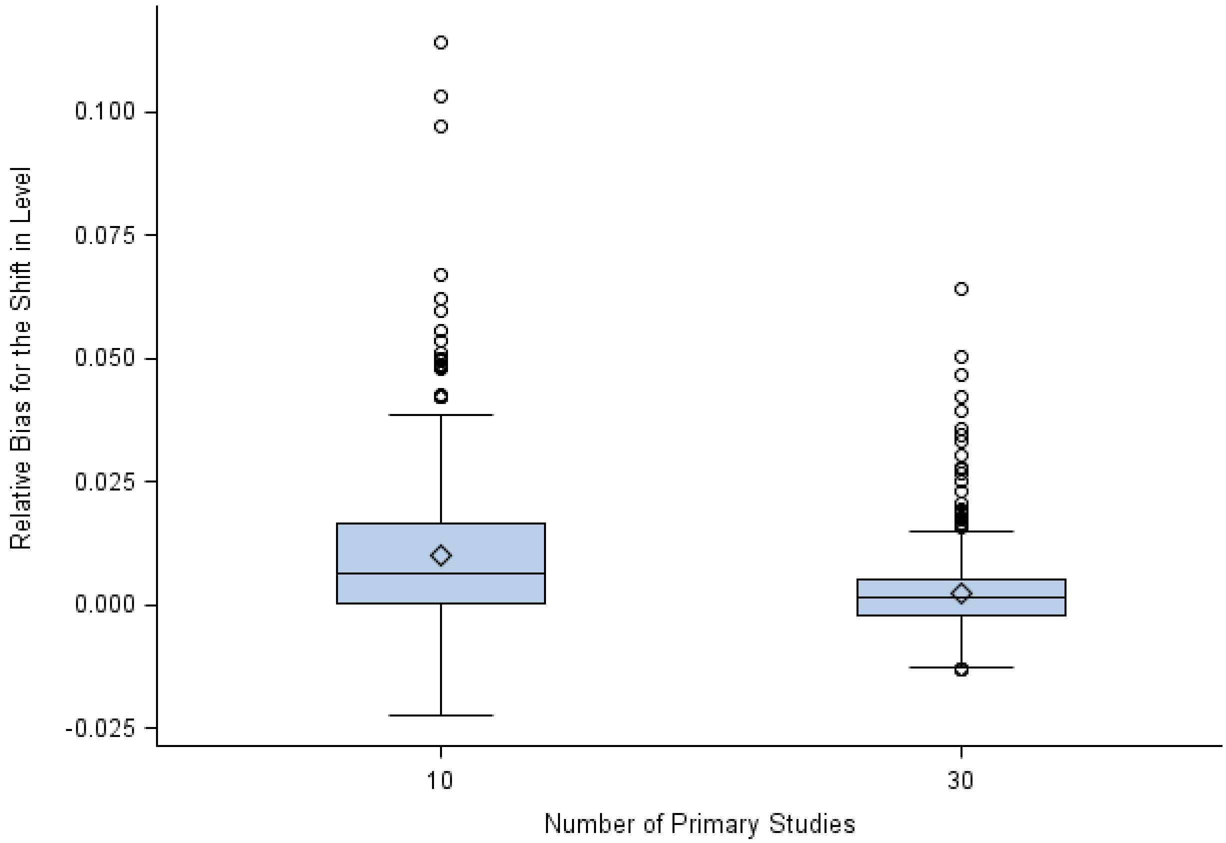

Figure A1. The relationship of the trimmed distribution for the level-three variance of the phase effect across the number of studies to be included in the meta-analysis.

Furthermore, the relative bias for the level-three variance of the phase effect (shift in level) tended to decrease as the variance was shifted from being mostly at level one (*M* = 0.01, *SD* = 0.02) to the upper levels (*M* =0.002, *SD* = 0.008). The variability also tended to decrease with more total variance, or again, as the variance was shifted from being mostly at level-one to being mostly at the upper levels.

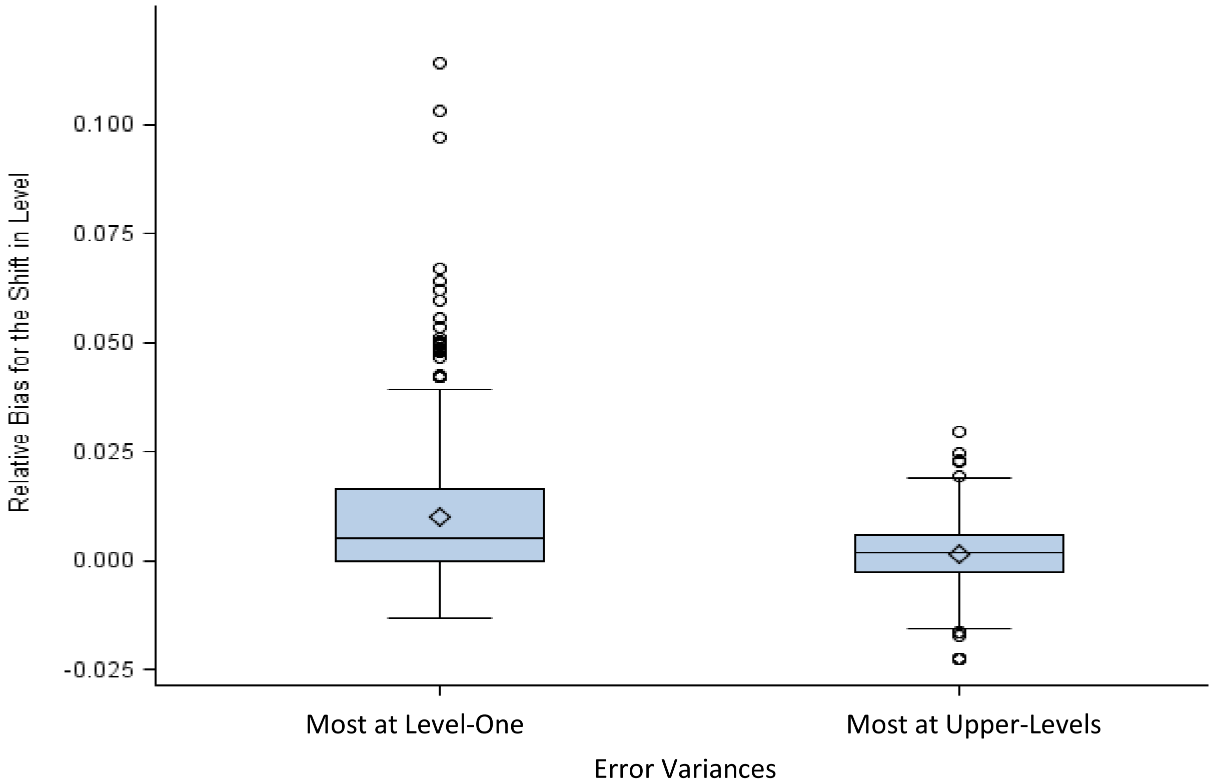

Figure A2. The relationship of the trimmed distribution for the level-3 variance of the phase effect across the variances of the error terms.

Level-three variance for the overall average treatment effect for the interaction effect (shift in slopes). The variance for the bias for the level-three variance of the interaction effect was further investigated for the trimmed relative bias. The GLM models, which included 5-way interactions and explained 93.7% of the variability, revealed two significant factors that the variances of the error terms ($\eta^2 = 0.06$) and the interaction of the type of model and the series length (0.07) had at least a medium effect. The mean relative bias decreased slightly as the variances shifted from being mostly at level one (*M* = 0.018, *SD* = 0.04) to most of the variance being at the upper levels (*M* = 0.004, *SD* = 0.012). The variability also tended to decrease as the variability shifted to being most at the upper levels as reflected below by Figure A3.

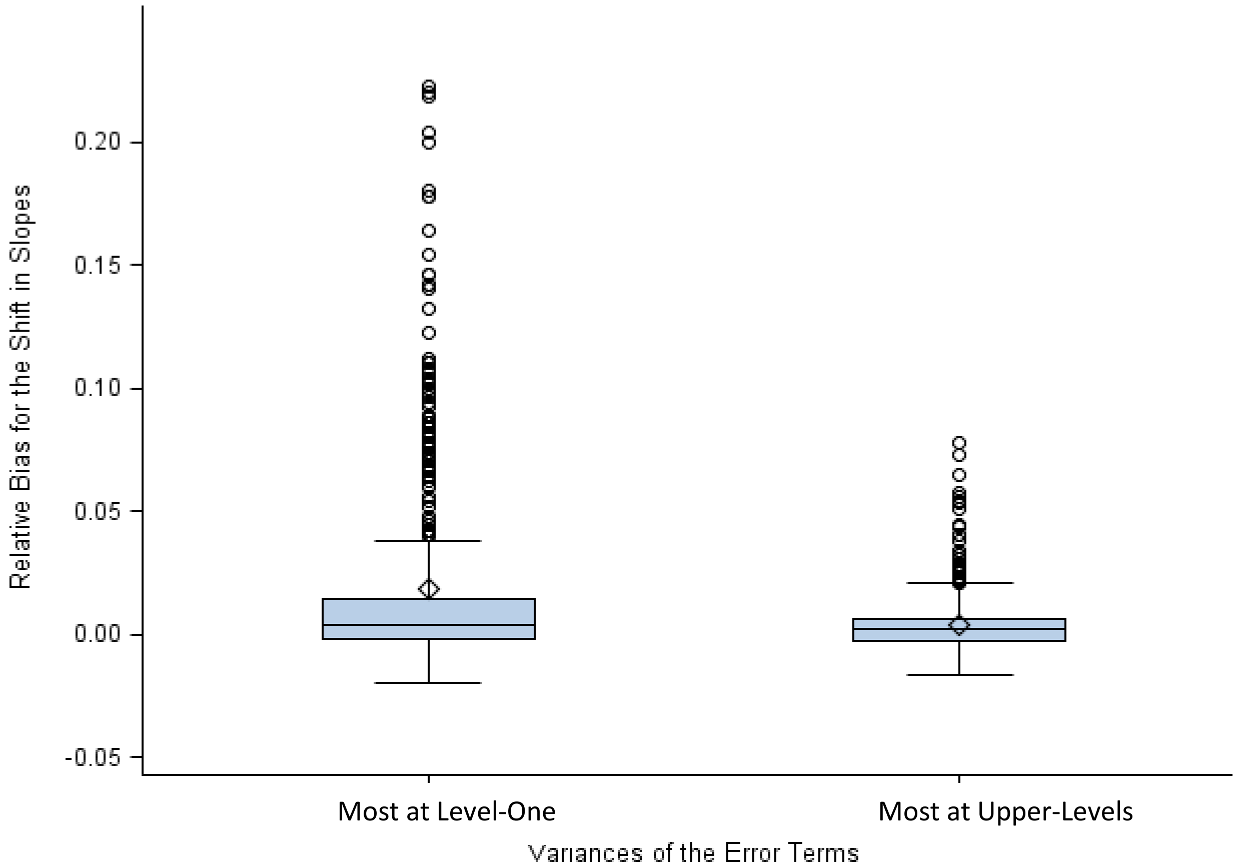

Figure A3. The bias for the trimmed distribution for the level-3 variance for the interaction effect (shift in slopes) as a function of the variance for the error terms.

The relationship of the bias for the level-three variance for the shift in slopes was then analyzed as a function of the interaction of the type of model and the series length. The line graph (see Figure A4 below) illustrates that the bias is higher across all seven models when there is a shorter series length, however this is accentuated for the ID model. The ID model shows a huge increase in the bias when the series length is 10.

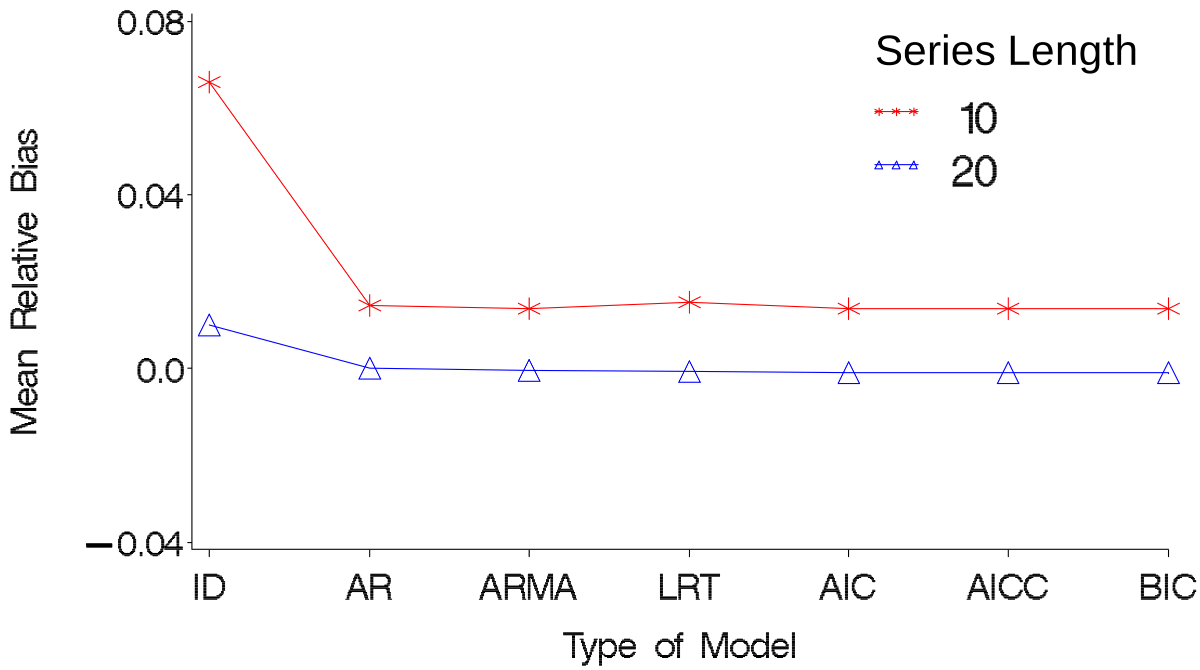

Figure A4. The relative bias for the level-3 variance for the shift in slopes for the interaction of the type of model and the series length.

Level two variance for the phase effect (shifts in level). The model, including 3-way interactions, explained 99.7% of the variability. The results revealed one medium or larger effect: the interaction of the level of autocorrelation and the type of model (η^2= 0.22). Graphs were produced to further examine the trimmed relative bias values.

Figure A8 below illustrates that the mean bias is similar across all of the models with the exception of the ID model. For the ID model, the mean relative bias tends to increase greatly as the level of autocorrelation increases.

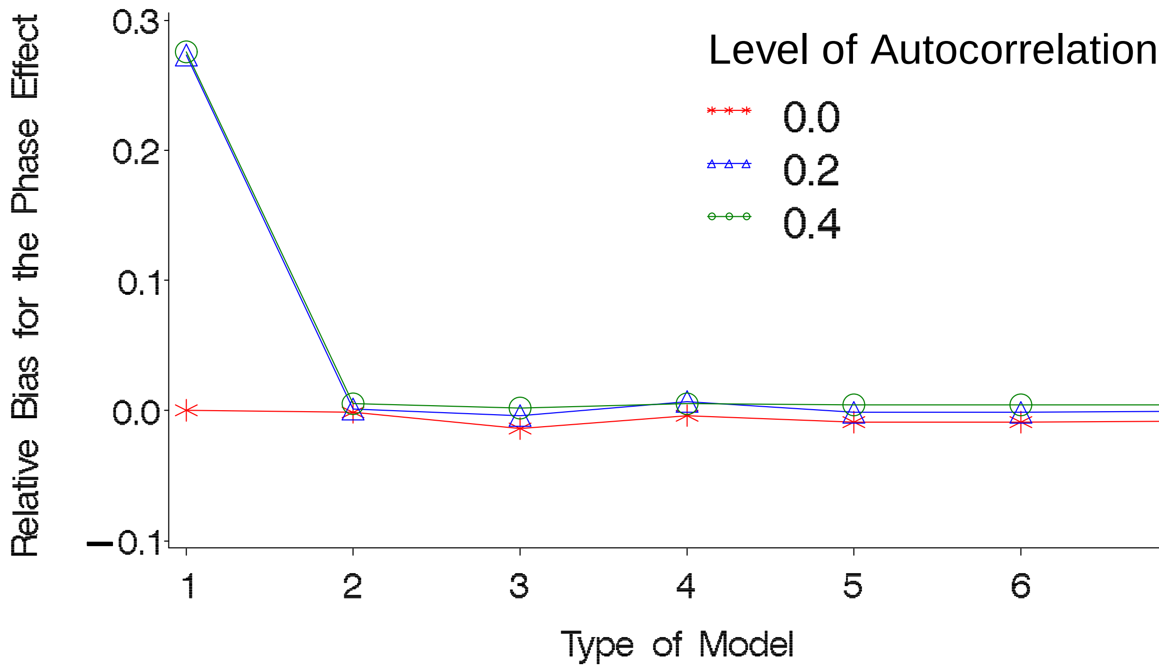

Figure A5. The distribution of the trimmed relative bias for the variance for the level-two phase effect as a function of the interaction between the level of the autocorrelation and the type of model.

Level two variance for the phase effect (shift in slope). The model, including 4th order interactions explained 97.7% of the variability and revealed two medium or larger effects: the three-way interaction of the series length and the variances of the error terms and the level of autocorrelation (η^2= 0.06) and the two-way interaction of the level of autocorrelation and the type of model (η^2= 0.19). The graphs below further examine the relationship of both effects on the mean relative bias for the level-two variance for the shift in slopes.

Figure A6 (top panel) illustrates that for the shorter series length of 10 and when most of the variance is at level one, then mean bias increases as the level of autocorrelation increases. However, when the series length is shorter and most of the variance is at the upper levels

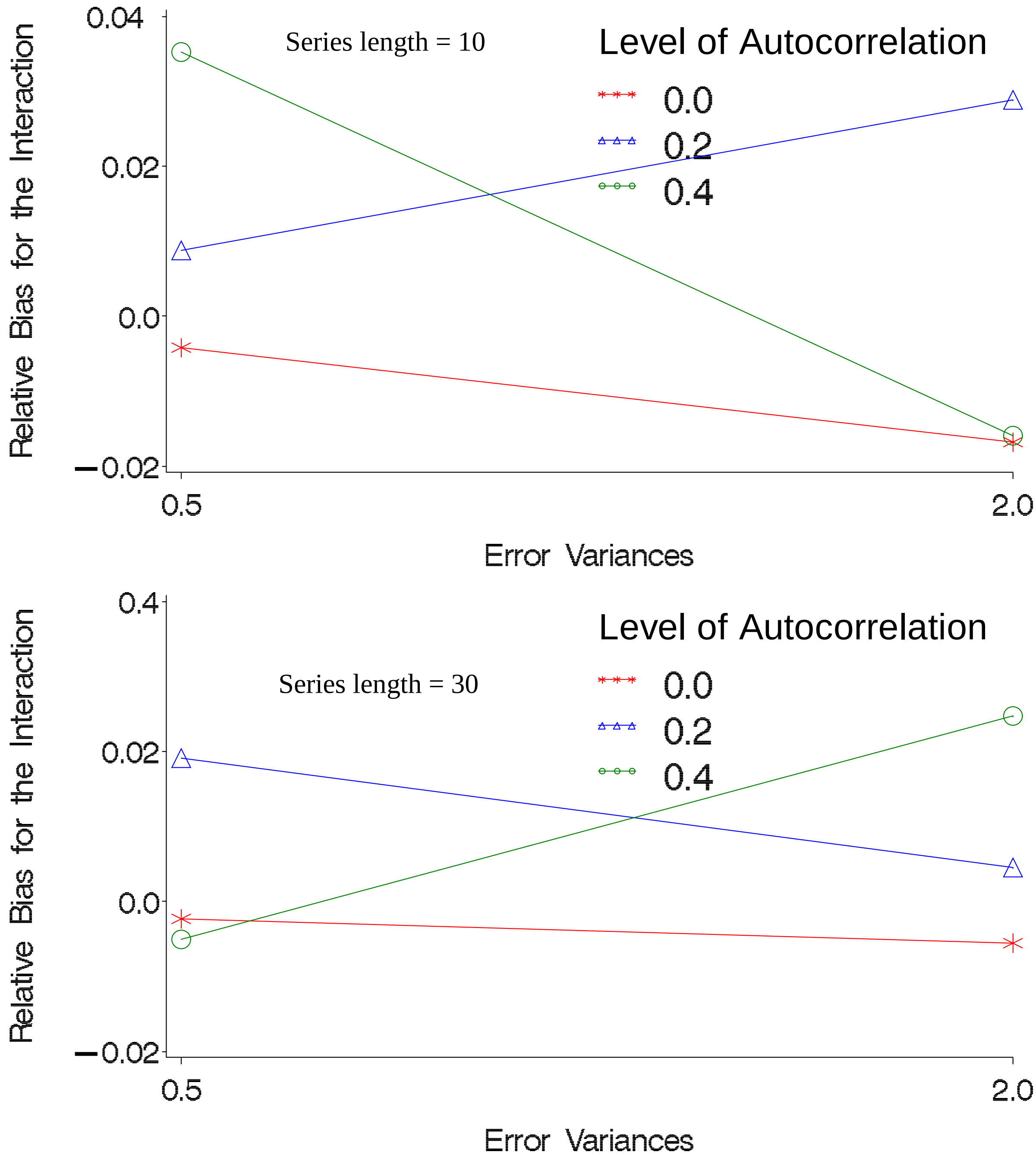

Figure A6. The effect on the mean relative bias for the level two variance for the shift in slopes as a function of the three way interaction between the level of autocorrelation, the error variances and the series length.

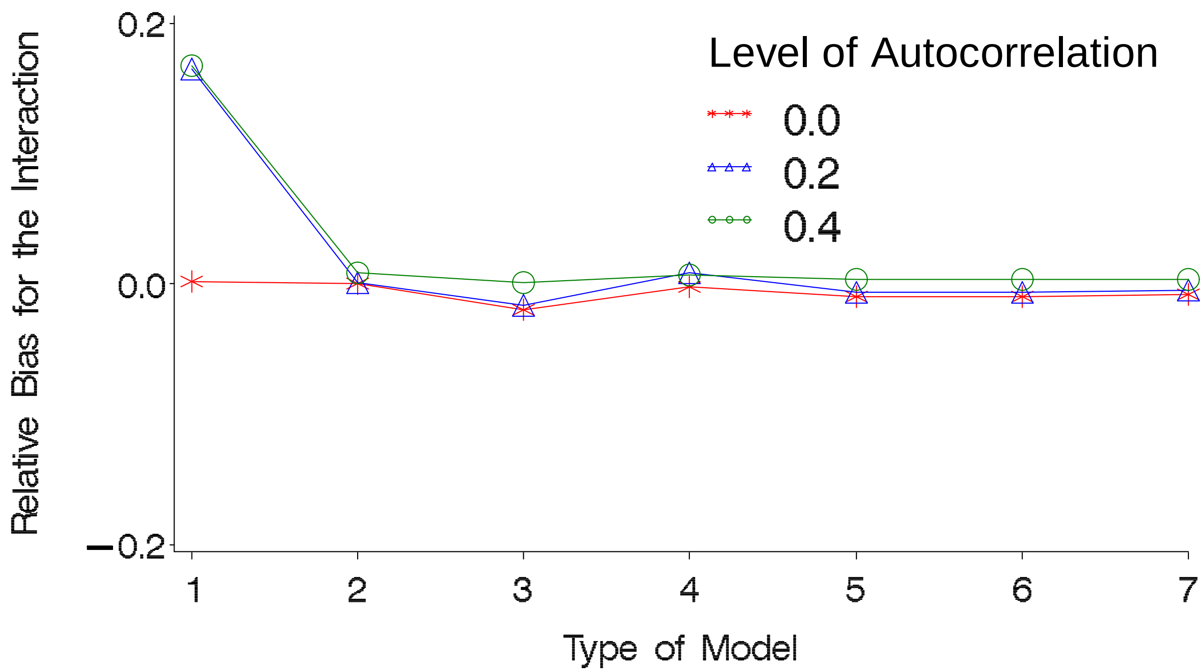

Figure A7. Distribution of trimmed relative bias for the shift in slopes as a function of the interaction between the level of autocorrelation and the type of model.

Figure A7 above displays little difference for the mean bias across the seven models with the exception of the ID model. Again, for the ID model, the mean bias increases as the level of autocorrelation increases.

Root Mean Square Error (RMSE)

The next section describes the additional analyses that were conducted for the trimmed RMSE values for the variance components for each of the treatment effects across both of the upper levels for the model.

Level-three variance for the overall average treatment effect for the phase effect (shift in level). The GLM models (including two-way interactions) explained 99% of the variability and revealed one medium or larger effect: the interaction of the number of primary studies with the variances of the error terms (η^2= 0.06). Further examination of the interaction with the RMSE values illustrated (see Figure A8) that for the shorter series length (10), the

RMSE values tend to be greater than when the series length is longer (30), however the gap is even greater when the variances of the error terms is mostly at the upper levels.

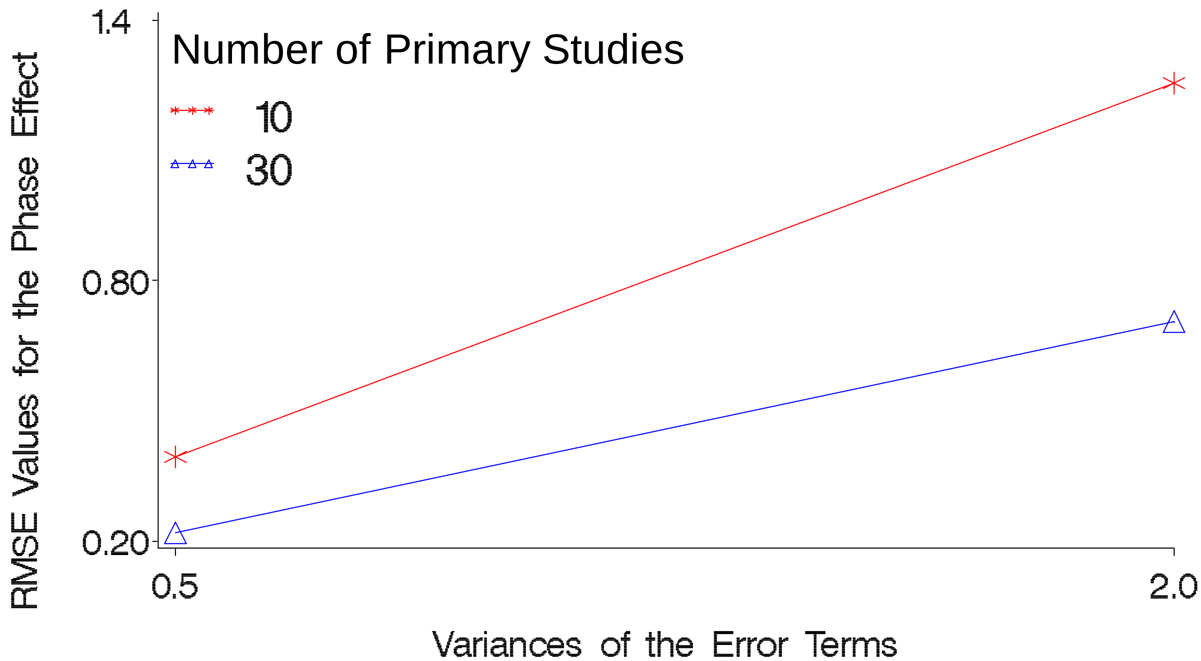

Figure A8. The line graph illustrates the effect of the interaction of the number of primary studies and the variances of the error terms on the trimmed RMSE values for the level-three variance for the shift in level.

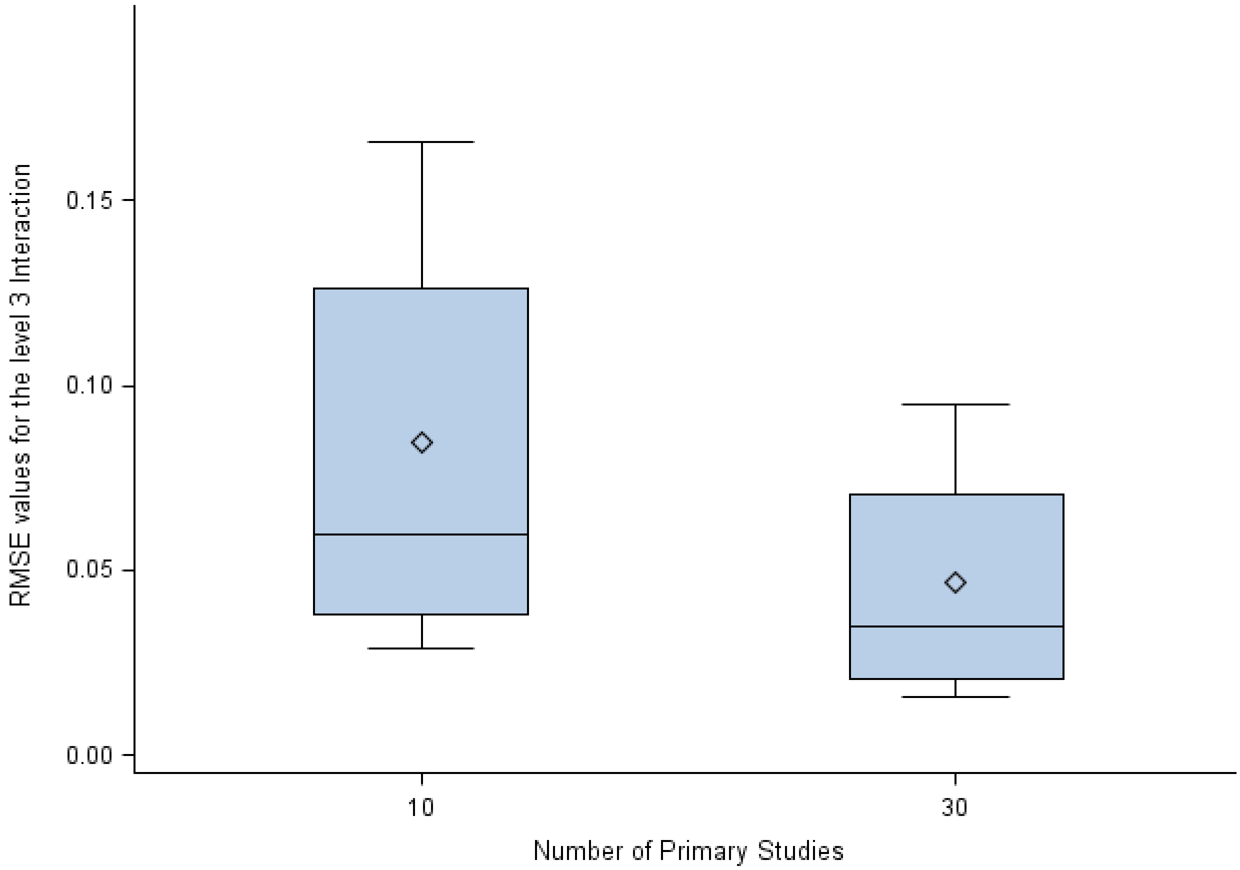

Figure A9. The relationship of the trimmed RMSE values for the level-three variance for the interaction effect across the number of primary studies.

Level-three variance for the overall average treatment effect for the interaction effect (shift in slopes). The GLM models included two-way interactions, which explained 99% of the variability found that there were two medium or larger effects. The number of primary studies (η^2= 0.20) and the variances of the error terms (η^2= 0.67) tended to have the most impact on the trimmed RMSE values for the level-three variance for the interaction effect. The two boxplots below, Figures A9 and A10, respectively, displays these effects and their impact on the RMSE values.

Figure A9 above revealed that the mean and the variability in the RMSE values tended to decrease with increased number of primary studies. Furthermore, in Figure A10 below, the mean and the variability of the RMSE values tended to increase as the variance was shifted from being mostly at level one to most of the variance being at the upper levels.

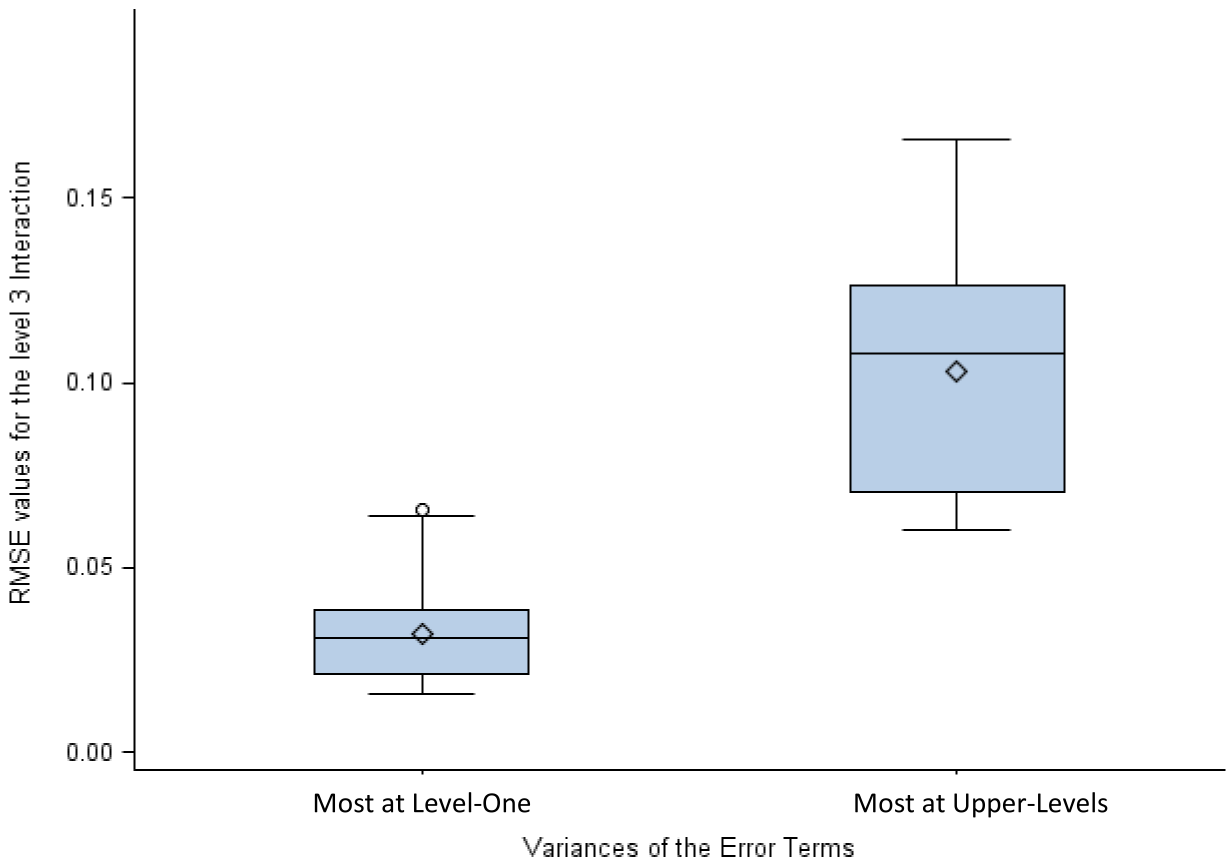

Figure A10. The relationship of the trimmed RMSE values for the level-three variance for the interaction effect across the error variances.

Level-two variance for the overall average treatment effect for the shift in level. The model, including third order interactions, explained 95.9% of the variability. There were three medium or larger effects: number of participants (η^2= 0.12), number of studies to be included in the meta-analysis (η^2= 0.22), the variance of the error terms (η^2= 0.43).

First, the mean for the RMSE values for the level-two variance for the shift in level decreased as the number of the participants increased from 4 to 8. The variability also tended to decrease with increased number of participants from 4 to 8 as is displayed below in Figure A11.

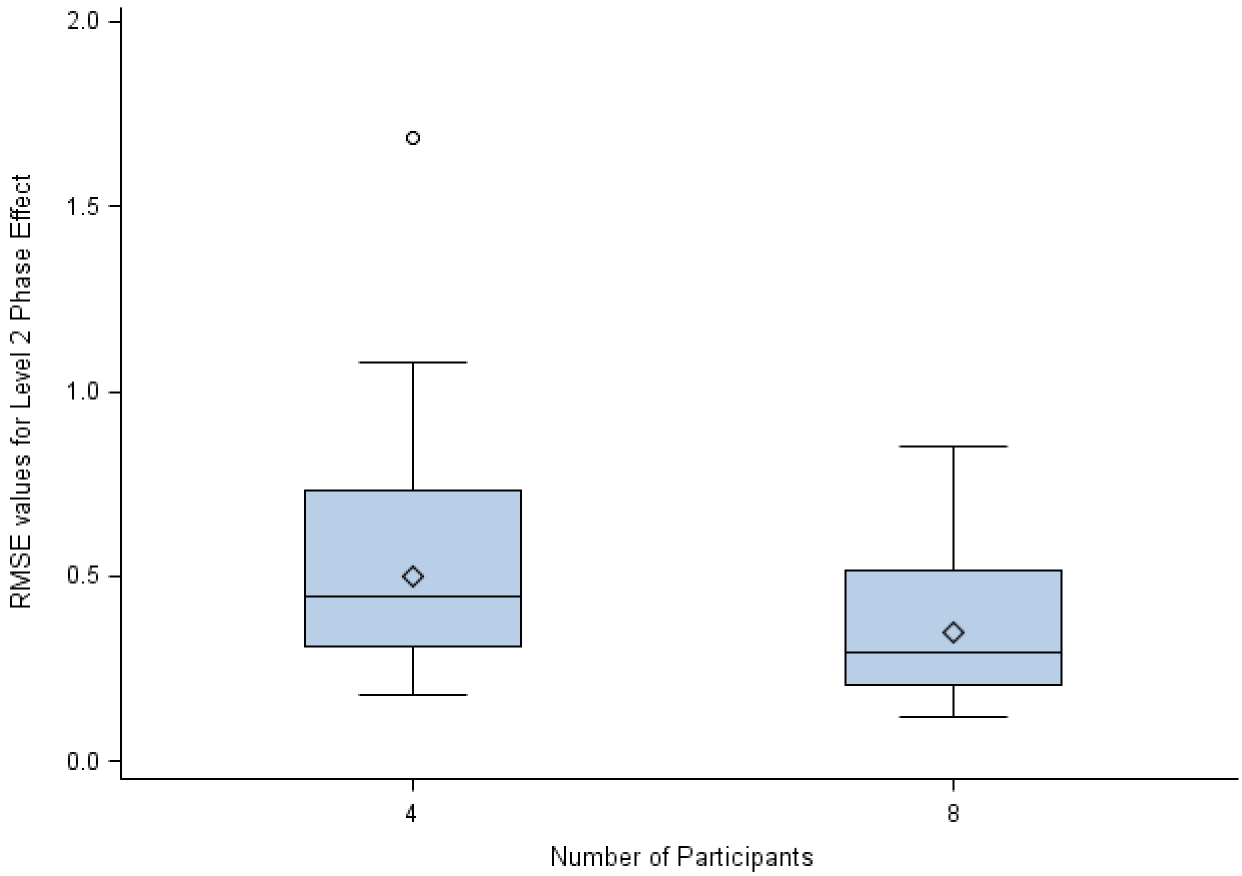

Figure A11. The box plot illustrating the distribution of the trimmed RMSE values for the variance for the phase effect as a function of the number of participants.

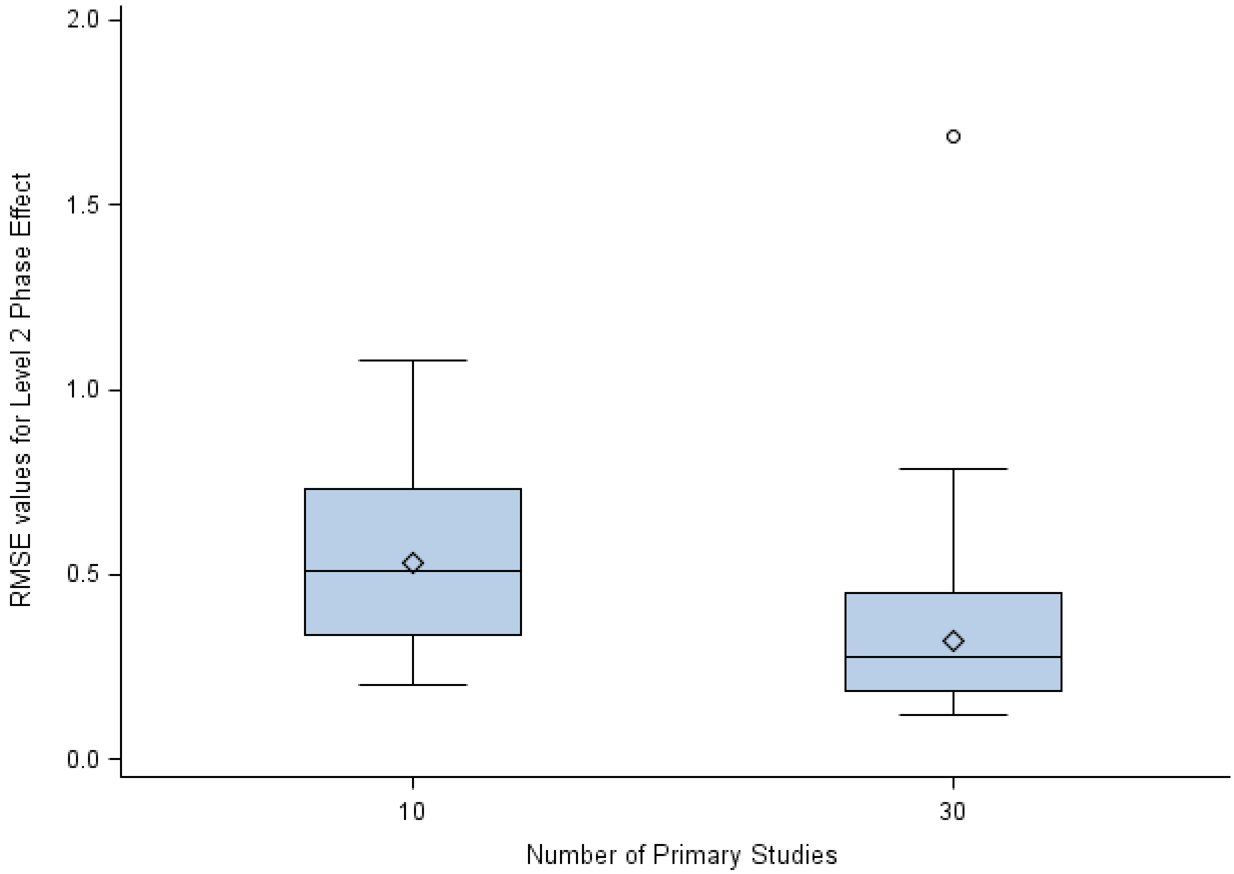

Figure A12. The box plot illustrating the distribution of the trimmed RMSE values for the variance for the phase effect as a function of the number of primary studies.

The figure above (Figure A12) shows the RMSE values for the level-two shift in level as a function of the number of primary studied. The mean and the variability of the RMSE values tended to decrease as the number of primary studies increased from 10 to 30. Furthermore, the variability appeared similar across the levels for the number of primary studies, with the exception of the outlier when the number of primary studies was larger (= 30).

Conversely, the mean RMSE values tended to increase as the variance of the error terms (see Figure A13 below) were shifted to mostly being at level one to mostly being at the upper levels. The variability also tended to increase, with the presence of an outlier, with increased total variability.

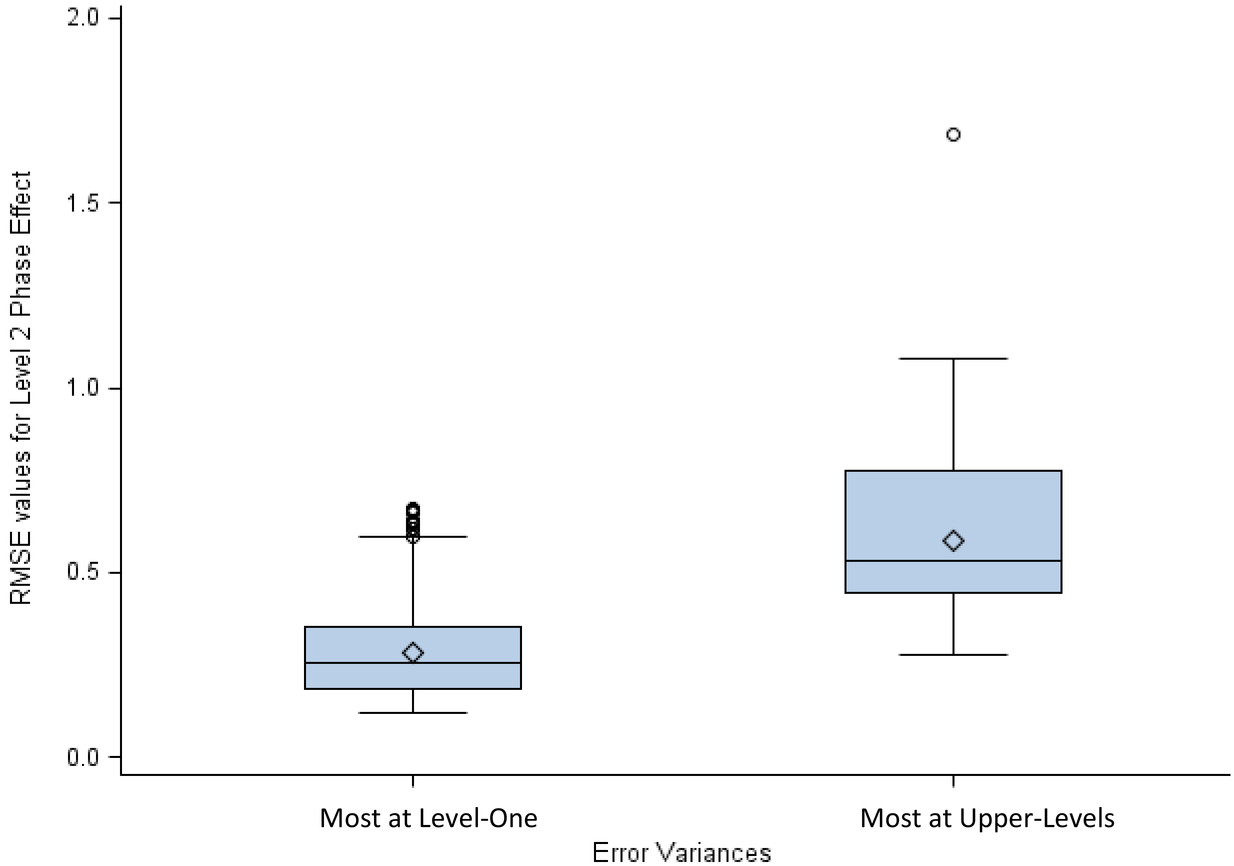

Figure A13. The box plot illustrating the distribution of the trimmed RMSE values for the variance for the phase effect as a function of the variance of the error terms.

Level-two variance for the overall average treatment effect for the interaction effect (shift in slopes). GLM models were run to further examine the relationship of the design factors with the trimmed level two variance for the shift in slopes. The model including two-way interactions explained 98.4% of the variability and revealed four medium or larger effects: number of participants (η^2= 0.08), the number of primary studies to be included in meta-analysis (η^2= 0.13), the series length (η^2= 0.26), and the variance of the error terms (η^2= 0.35).

As the number of participants (see Figure A14), the number of primary studies (see Figure A15), and the series length (see Figure A16) increased, then the mean RMSE values decreased. Furthermore, the mean RMSE values increased as the variance of the error terms (see Figure A17) were shifted from most of the variance being at level-one to most of the variance being at the upper levels. Both levels for the number of participants seemed to include several outliers at the upper ends.

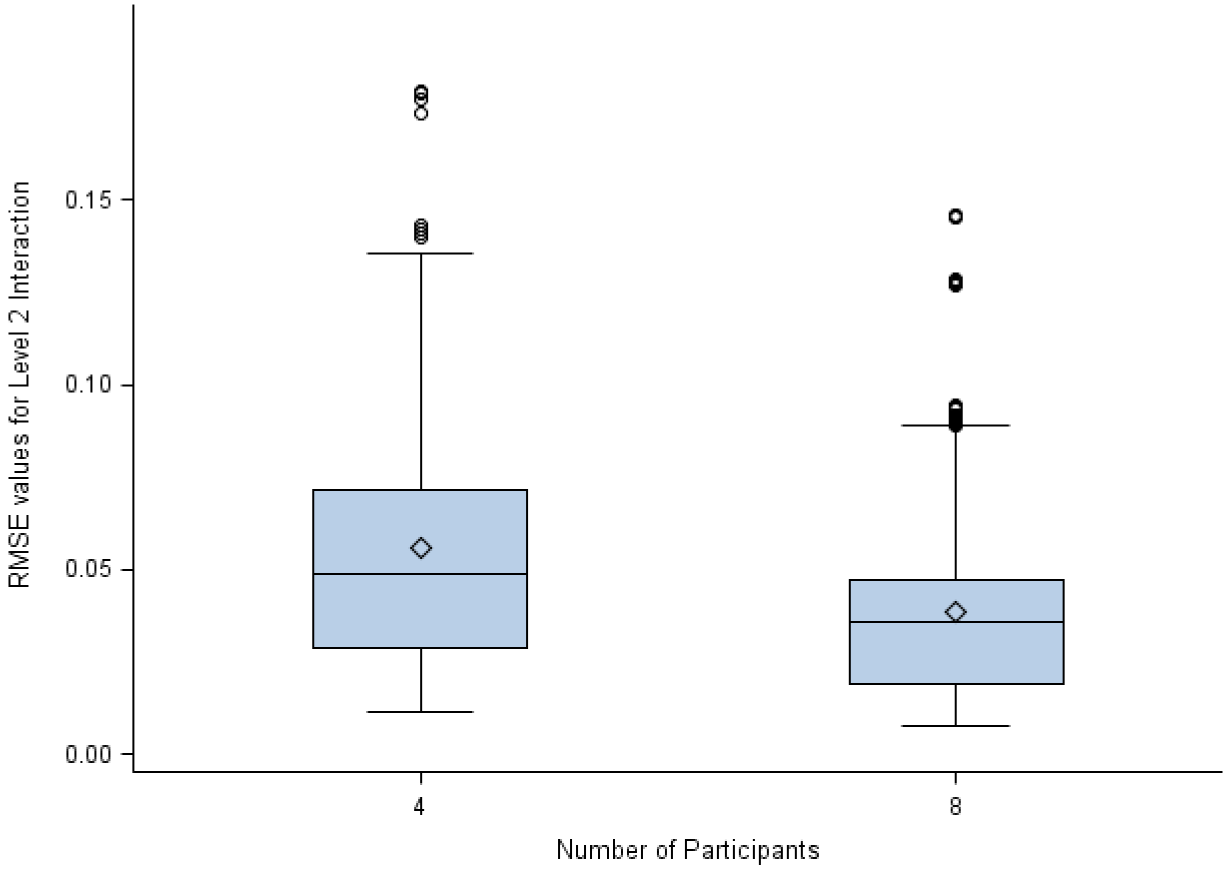

Figure A14. The box plot illustrating the distribution of the trimmed RMSE values for the variance (level two) for the shift in slopes as a function of the number of participants.

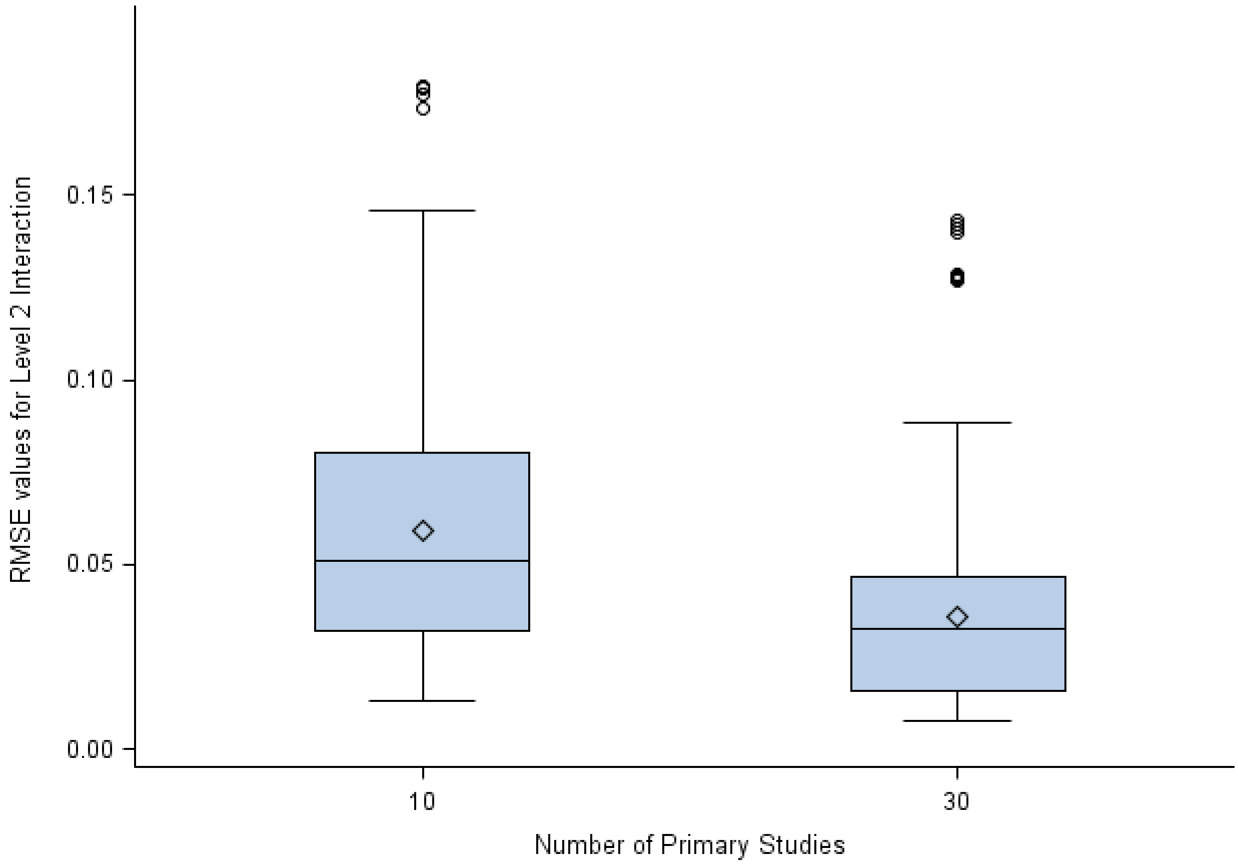

Figure A15. The box plot illustrating the distribution of the trimmed RMSE values for the variance (level two) for the shift in slopes as a function of the number of primary studies to be included in meta-analysis.

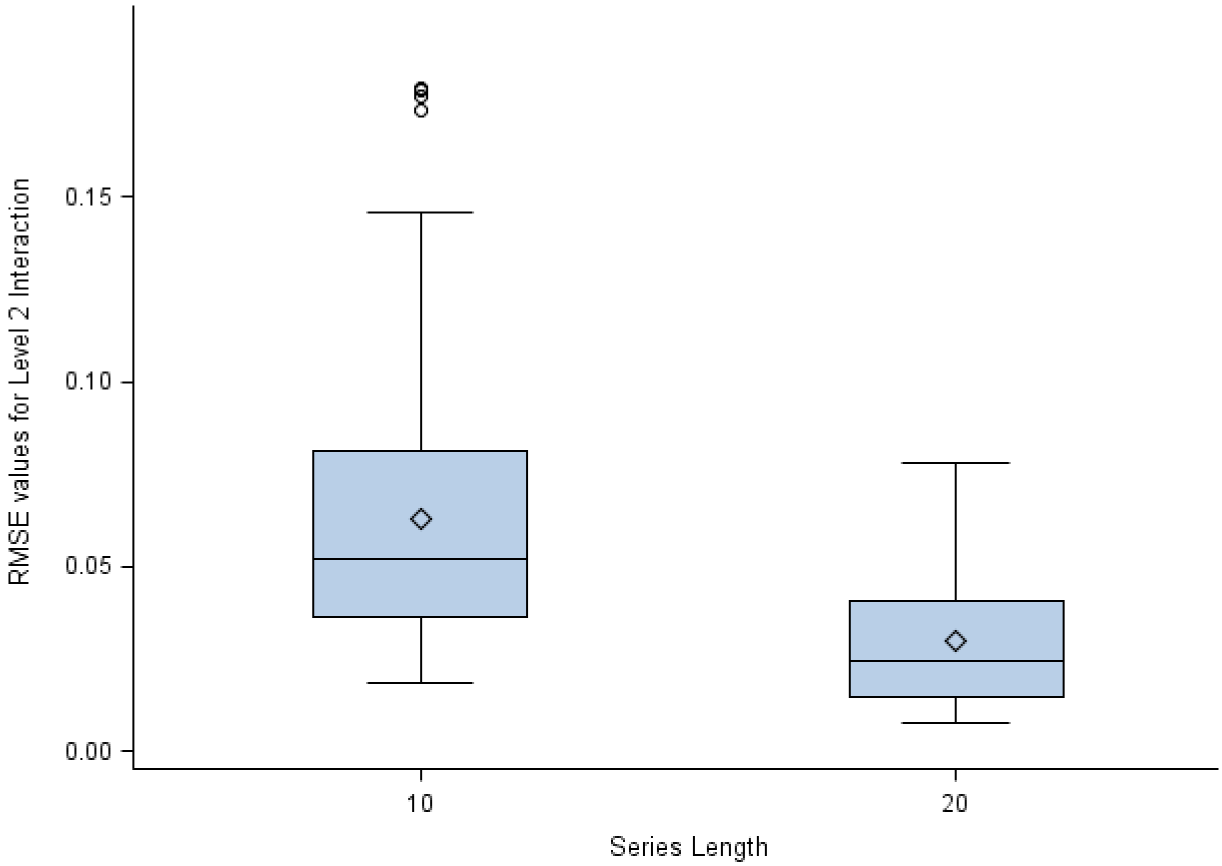

Figure A16. The box plot illustrating the distribution of the trimmed RMSE values for the variance (level two) for the shift in slopes as a function of the series length.

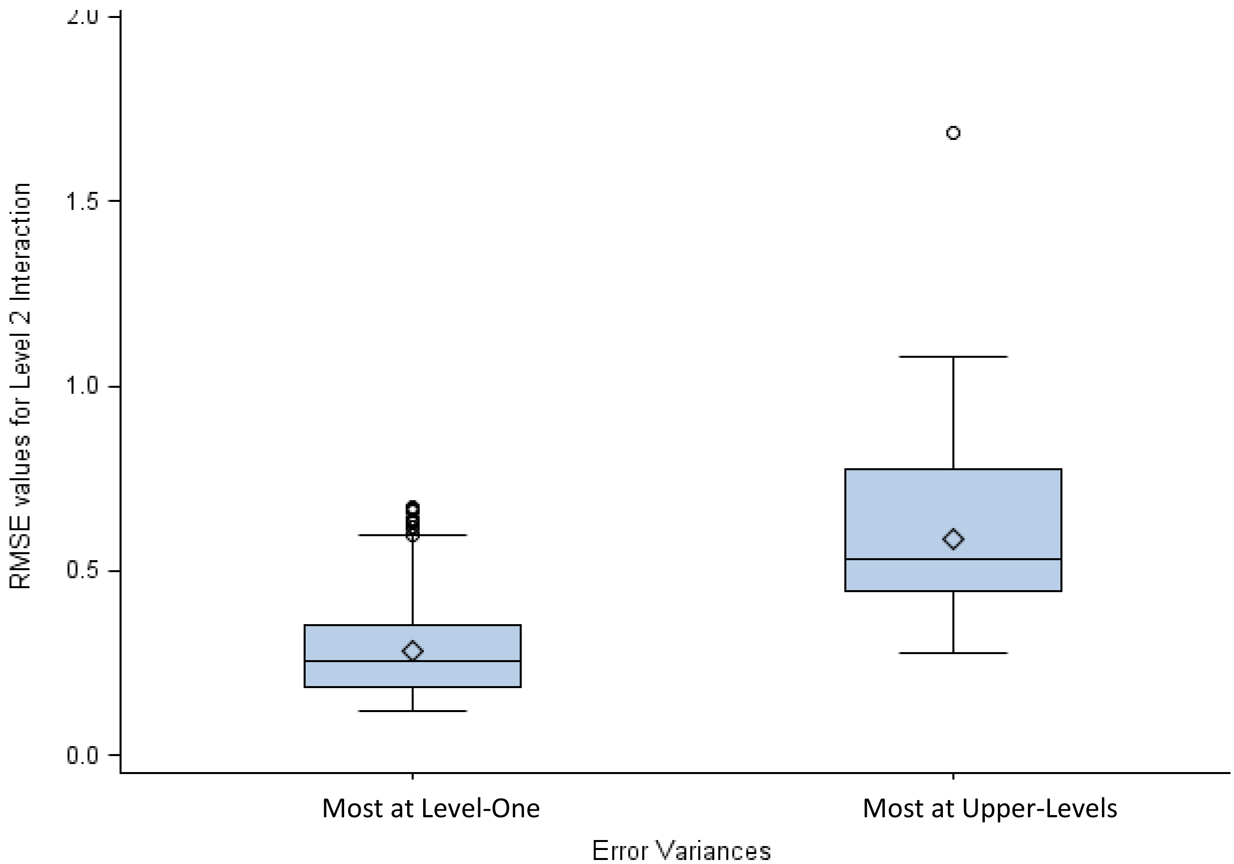

Figure A17. The box plot illustrating the distribution of the trimmed RMSE values for the variance (level two) for the shift in slopes as a function of the error variances.

Level one variance or residual variance. The resulting model included 3-way interactions and explained 98% of the total variability. The following medium effects were found: the series length or number of observations ($\eta^2 = 0.06$), the interaction of the level of the autocorrelation parameter and the type of model ($\eta^2 = 0.07$), the interaction of the variances of the error terms and the type of model ($\eta^2 = 0.11$). Graphs were created to analyze the association of the mean RMSE values for the level one variance with these effects. The graph below (see Figure A18) depicts the association of the trimmed RMSE values for the level one variance and the number of observations (series length). Specifically, the graph demonstrates that as the series length is increased from 10 then the mean RMSE for the level one variance is decreased.

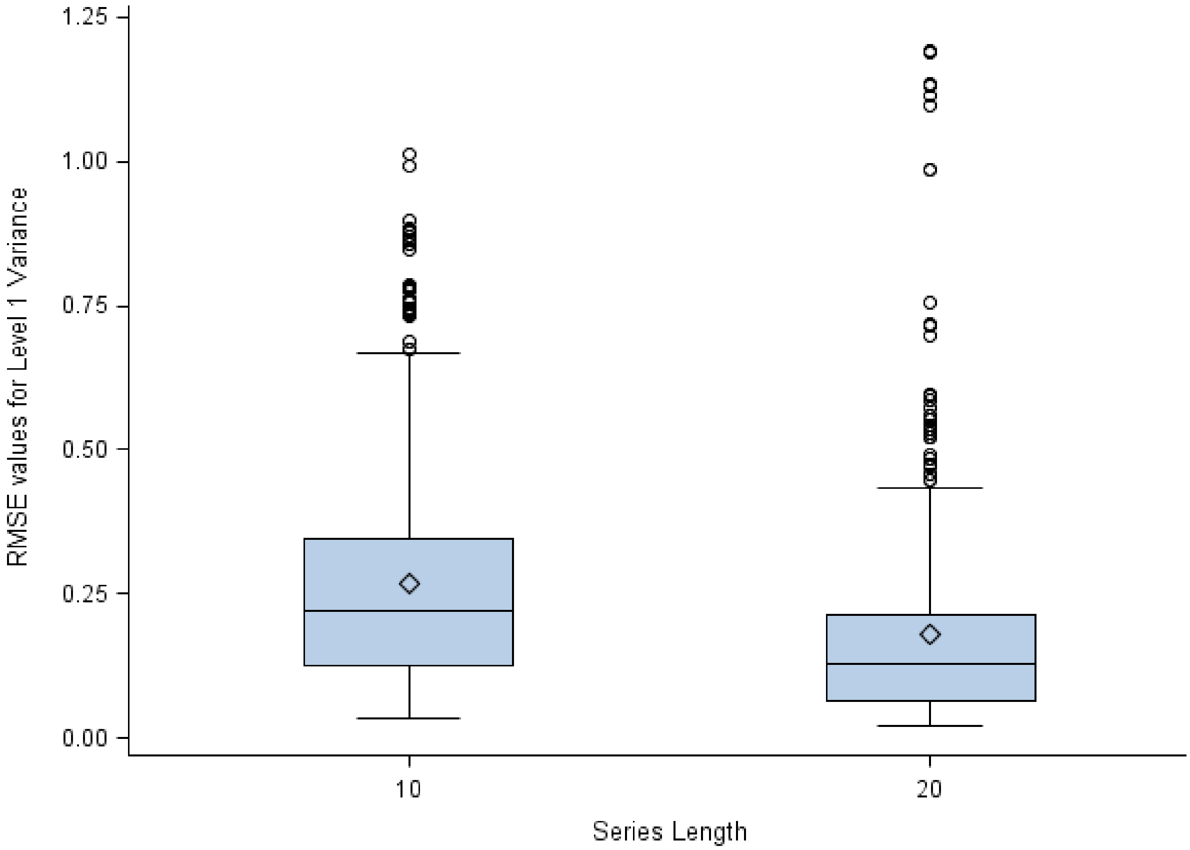

Figure A18. The box plot illustrating the distribution of the trimmed RMSE values for the level one variance across the series length.

The line graph below in Figure A19 illustrates the relationship of the mean RMSE values for the level one variance and the interaction of the variances of the error terms and the type of model. Furthermore, the graph shows that when most of the variance is at level one the mean RMSE values is consistently lower across all of the models than when most of the variance is at the upper levels. However, that difference is greater when most of the variance is at the upper levels for some of the models. For example, the ARMA(1,1) model has the greatest increase in the mean RMSE values for the level one variance when the variances of the error terms are shifted from most of the variance being at level one to most of the variance being at the upper levels. The smallest difference is seen for the ID model, in which there was very little change in the mean RMSE values for the level one variance when the variances of the error terms were shifted from most of the variance being at level one to most of the variance being at the upper levels.

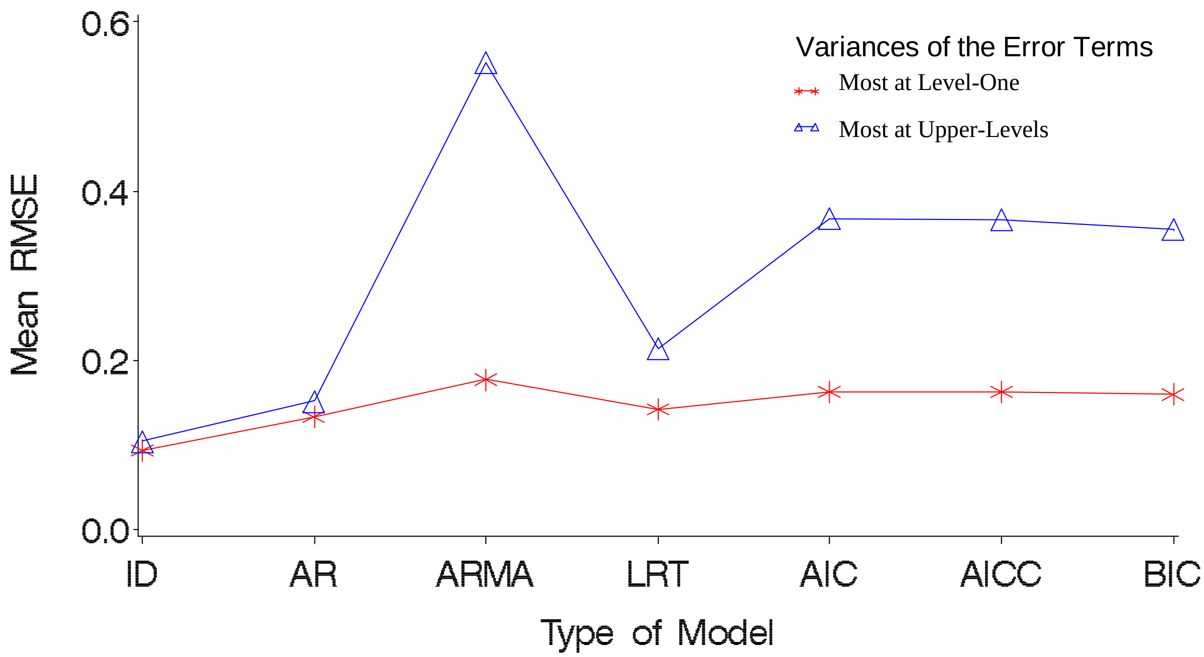

Figure A19. Line graph illustrating the relationship between the mean RMSE values for the level one variance and the interaction effect of the variances of the error terms and type of model.

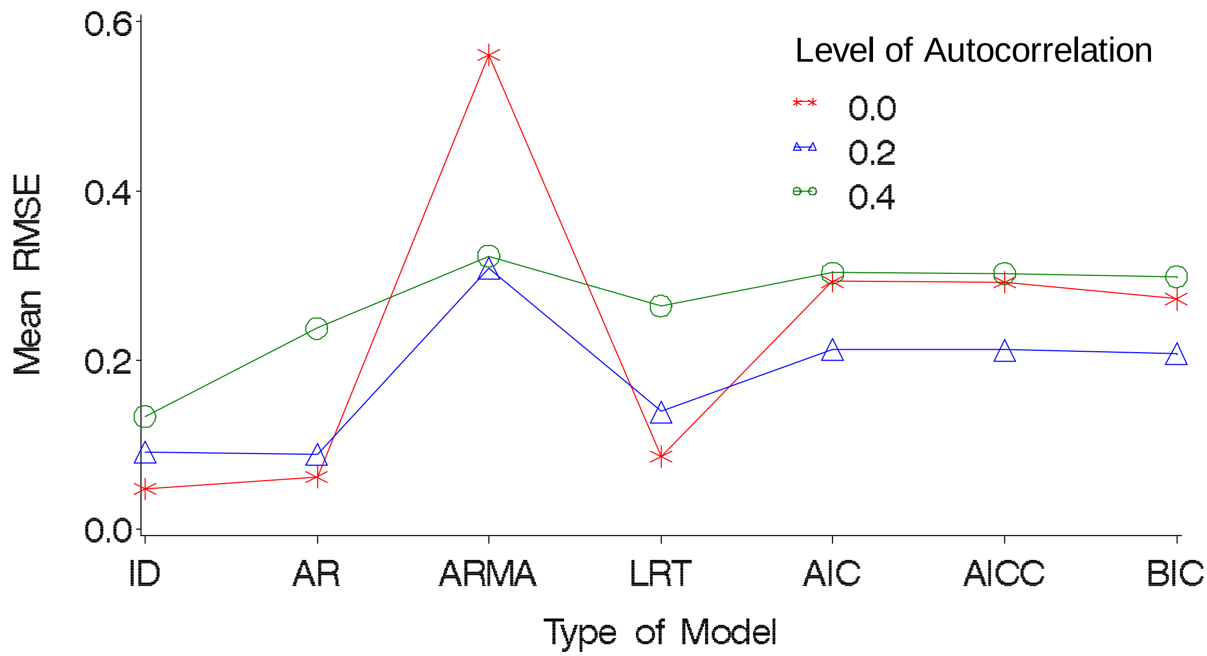

Figure A20. Line graph illustrating the association between the mean RMSE values for the level one variance and the interaction effect of the level of the autocorrelation parameter and type of model.

Lastly, the association of the mean RMSE values for the level one variance and the interaction effect of the level of the autocorrelation parameter and the type of model were analyzed using a line graph (see Figure A20). Across some of the models, such as the ID, AR(1), and models selected by LRT, the mean RMSE values for the level one variance tended to increase as the level of the autocorrelation parameter increased. For the remainder of the fit-index selected models (AIC, AICC, and BIC), the mean RMSE values is smallest when $\rho = 0.2$ and largest when $\rho = 0.4$. For example, for the models selected by BIC, the mean RMSE values for the level one variance are as follows: $\rho = 0.2$ (*M* = 0.21, *SD* = 0.17); $\rho = 0.0$ (*M* = 0.27, *SD* = 0.19); $\rho = 0.4$ (*M* = 0.30, *SD* = 0.14). Similar patterns are observed for the AIC and AICC models. Finally, a vast decline in the mean RMSE values is noticed for the ARMA(1,1) model as the level of the autocorrelation is increased from 0.0 (*M* = 0.56, *SD* = 0.43) to 0.2 (*M* = 0.31, *SD* = 0.25) to 0.4 (*M* = 0.32, *SD* = 0.18).

APPENDIX B: TABLES OF ETA-SQUARED VALUES

Table A1

Eta-Squared Values (η^2)for the Association of the Design Factors with the Proportion Correct for the ID Model

	η^2
Type of Fit Index	.771
Number of Primary Studies	.068
Number of Primary Studies*Type of Fit Index	.064
Error Variance	.038
Series Length	.027
Number of Participants	.006
Number of Primary Studies * Error Variance	.004
Number of Participants* Error Variance	.002
Series Length* Number of Primary Studies	.002
Level of the Fixed Levels	.002
Series Length* Error Variance	.001
Number of Participants * Type of Fit Index	.001
Series Length * Number of Participants	.001
Number of Participants *Level of Fixed Level	.0003
Error Variance * Type of Fit Index	.0003
Series Length * Type of Fit Index	.0001
Error Variance * Level of the Fixed Levels	.00007
Level of the Fixed Levels * Type of Fit Index	.00006
Number of Primary Studies * Level of the Fixed Levels	.00002
Number of Participants * Number of Primary Studies	.00001
Series Length * Level of the Fixed Levels	.00000
Total Explained	0.988

Table A2

Eta-Squared Values (η^2) for the Association of the Design Factors with the Proportion Correct for the AR(1) Model

	η^2
Type of Model	0.23552
Number of Primary Studies	0.15765
Series Length	0.09592
Autocorrelation	0.09292
Number of Participants	0.05866
Series Length*Number of Primary Studies	0.0507
Series Length*Autocorrelation	0.04417
Number of Primary Studies*Autocorrelation	0.03982
Series Length*Number of Participants	0.03029
Series Length*Number of Primary Studies*Autocorrelation	0.02755
Number of Primary Studies*Type of Model	0.02699
Number of Participants*Number of Primary Studies	0.02216
Number of Participants*Autocorrelation	0.02103
Series Length*Number of Participants*Autocorrelation	0.01305
Series Length*Number of Participants*Number of Primary Studies	0.0122
Number of Participants*Number of Primary Studies*Autocorrelation	0.01108
Error Variance	0.00712
Error Variance*Autocorrelation	0.00683
Autocorrelation*Type of Model	0.00598
Number of Primary Studies*Autocorrelation*Type of Model	0.00449
Series Length*Autocorrelation*Type of Model	0.00448
Series Length*Error Variance	0.00203
Number of Participants*Autocorrelation*Type of Model	0.00199
Series Length*Error Variance*Autocorrelation	0.00168
Error Variance*Type of Model	0.0011
Series Length*Number of Primary Studies*Type of Model	0.00109
Series Length*Type of Model	0.00077
Series Length*Number of Participants*Type of Model	0.00063
Number of Participants*Number of Primary Studies*Type of Model	0.0005
Series Length*Number of Participants*Error Variance	0.00035
Series Length*Number of Primary Studies*Error Variance	0.0003
Number of Participants*Type of Model	0.00026
Series Length*Error Variance*Type of Model	0.00024
Number of Primary Studies*Error Variance*Type of Model	0.00023
Number of Primary Studies*Error Variance	0.00019
Error Variance*Autocorrelation*Type of Model	0.00016
Number of Participants*Number of Primary Studies*Error Variance	0.00015
Number of Primary Studies*Error Variance*Fixed Level	0.00014
Level of Fixed Level	0.00009

Table A2 (Continued)

	η^2
Number of Participants*Level of Fixed Level	0.00009
Number of Participants*Error Variance*Type of Model	0.00009
Number of Participants*Error Variance*Autocorrelation	0.00008
Series Length*Number of Participants*Fixed Level	0.00007
Number of Primary Studies*Error Variance*Autocorrelation	0.00007
Level of Fixed Level*Autocorrelation	0.00007
Number of Participants*Number of Primary Studies*Fixed Level	0.00005
Error Variance*Level of Fixed Level*Autocorrelation	0.00005
Number of Primary Studies*Level of Fixed Level*Autocorrelation	0.00003
Number of Participants*Error Variance*Fixed Level	0.00002
Series Length*Level of Fixed Level	0.00002
Series Length*Error Variance*Fixed Level	0.00002
Series Length*Number of Primary Studies*Fixed Level	1.49E-05
Number of Participants*Level of Fixed Level*Autocorrelation	1.29E-05
Series Length*Level of Fixed Level*Autocorrelation	1.14E-05
Error Variance*Level of Fixed Level	8.09E-06
Error Variance*Fixed Level*Type of Model	7.07E-06
Number of Participants*Fixed Level*Type of Model	6.78E-06
Level of Fixed Level*Type of Model	5.19E-06
Level of Fixed Level*Autocorrelation*Type of Model	2.99E-06
Number of Primary Studies*Fixed Level*Type of Model	1.6E-06
Series Length*Fixed Level*Type of Model	1.52E-06
Number of Participants*Error Variance	1.39E-06
Number of Primary Studies*Level of Fixed Level	7.88E-07

Table A3

Eta-Squared Values (η^2)for the Association of the Design Factors with the Proportion Correct for the ARMA(1,1) Model

	η^2
Type of Model	0.71737
Number of Primary Studies	0.09608
Number of Primary Studies*Type of Model	0.07899
Autocorrelation	0.02439
Series Length	0.02074
Number of Participants	0.0145
Series Length*Type of Model	0.01005
Error Variance	0.00531
Series Length*Number of Primary Studies	0.00451
Error Variance*Autocorrelation	0.00347
Number of Participants*Type of Model	0.00285

Table A3 (Continued)

	η^2
Series Length*Number of Participants	0.00271
Number of Participants*Number of Primary Studies	0.00145
Number of Primary Studies*Error Variance	0.00078
Error Variance*Type of Model	0.0003
Series Length*Error Variance	0.00021
Number of Participants*Error Variance	0.00017
Autocorrelation*Type of Model	0.00013
Number of Primary Studies*Fixed Level	0.00007
Series Length*Autocorrelation	0.00005
Number of Primary Studies*Autocorrelation	0.00005
Fixed Level	0.00002
Error Variance*Fixed Level	0.00002
Fixed Level*Type of Model	0.00002
Number of Participants*Fixed Level	0.00001
Series Length*Fixed Level	0.00001
Number of Participants*Autocorrelation	0.00001
Fixed Level*Autocorrelation	0
Moving Average	0
Series Length*Moving Average	0
Number of Participants*Moving Average	0
Number of Primary Studies*Moving Average	0
Error Variance*Moving Average	0
Autocorrelation*Moving Average	0
Fixed Level*Moving Average	0
Moving Average*Type of Model	0
Total Explained	

Table A4

Eta-Squared Values (η^2)for the Association of the Design Factors with the RMSE Values for the Shift in Level

	η^2
Error Variance	0.4921
Number of Primary Studies	0.4523
Number of Participants	0.01536
Series Length	0.00204
Autocorrelation	0.0003
Type of Model	0.00004
Moving Average	0.00002
Fixed Level	0
Total Explained	0.9622

Table A5

Eta-Squared Values (η^2)for the Association of the Design Factors with the RMSE Values for the Shift in Slope

	η^2
Error Variance	0.45497
Number of Primary Studies	0.44288
Series Length	0.03721
Number of Participants	0.02018
Autocorrelation	0.00148
Moving Average	0.0000
Type of Model	0.0000
Fixed Level	0.0000
Total Explained	0.9567

TableA6

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI Width for the Shift in Level

	η^2
Autocorrelation*Type of Model	0.1556
Error Variance	0.1044
Number of Primary Studies	0.1036
Type of Model	0.0753
Error Variance*Autocorrelation*Type of Model	0.0453
Error Variance*Type of Model	0.0257
Autocorrelation	0.0240
Number of Participants*fix*Error Variance*Autocorrelation*Type of Model	0.0177
Series Length*Number of Participants*Error Variance*Autocorrelation*Type of Model	0.0172
Series Length*Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.0156
Series Length*Autocorrelation*Type of Model	0.0134
Number of Participants*Autocorrelation*Type of Model	0.0132
Fixed Level*Error Variance*Autocorrelation*Type of Model	0.0123
Number of Participants*Type of Model	0.0116
Number of Primary Studies*Error Variance	0.0115
Series Length*Error Variance*Autocorrelation*Type of Model	0.0107
Number of Primary Studies*Type of Model	0.0106
Number of Primary Studies*Error Variance*Moving Average*Type of Model	0.0089
Series Length*Number of Primary Studies*Error Variance*the*Type of Model	0.0089
Number of Participants*Moving Average*Type of Model	0.0083
Series Length*Error Variance*Moving Average*Type of Model	0.0083
Error Variance*Moving Average*Type of Model	0.0082
Series Length*Number of Participants*Moving Average*Type of Model	0.0082
Series Length*Number of Participants*Number of Primary Studies*the*Type of Model	0.0082

Table A6 (Continued)

	η^2
Number of Participants*Number of Primary Studies*Moving Average*Type of Model	0.0081
Number of Primary Studies*fix*Error Variance*Autocorrelation*Type of Model	0.0080
Number of Participants	0.0080
Error Variance*Autocorrelation	0.0079
Number of Primary Studies*Autocorrelation*Type of Model	0.0074
Series Length*Number of Primary Studies*Autocorrelation*Type of Model	0.0073
Number of Participants*Fixed Level*Error Variance*Type of Model	0.0072
Number of Participants*Error Variance*Type of Model_	0.0069
Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.0049
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.0048
Number of Participants*Error Variance*Autocorrelation*Type of Model	0.0048
Number of Primary Studies*Error Variance*Type of Model_	0.0048
Series Length*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.0047
Series Length*Number of Primary Studies*Moving Average*Type of Model	0.0043
Number of Primary Studies*Moving Average*Type of Model_c	0.0042
Series Length*Number of Participants*Error Variance*the*Type of Model	0.0041
Number of Participants*Error Variance*Moving Average*Type of Model	0.0041
Number of Participants*Number of Primary Studies*Type of Model_	0.0041
Number of Participants*Number of Primary Studies*Error Variance*the*Type of Model	0.0040
Moving Average*Type of Model	0.0038
Series Length*Moving Average*Type of Model_c	0.0038
Series Length*Number of Participants*Type of Model_	0.0037
Number of Participants*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.0037
Number of Participants*Number of Primary Studies*fix*Error Variance*Type of Model	0.0037
Series Length*Fixed Level*Autocorrelation*Type of Model	0.0034
Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.0032
Fixed Level*Error Variance*Type of Model_	0.0030
Number of Participants*Fixed Level*Error Variance*Autocorrelation	0.0029
Series Length*Number of Participants*Error Variance*Autocorrelation	0.0029
Series Length*fix*Error Variance*Autocorrelation*Type of Model	0.0028
Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.0027
Series Length*Number of Primary Studies*Type of Model_	0.0027
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.0026
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.0026
Series Length*Number of Primary Studies*Error Variance*Autocorrelation	0.0026
Series Length*Number of Participants*Autocorrelation*Type of Model	0.0025
Series Length*Number of Participants*Number of Primary Studies*Type of Model	0.0024
Series Length*Autocorrelation	0.0022

Table A6 (Continued)

	η^2
Fixed Level*Error Variance*Autocorrelation	0.0020
Number of Participants*Autocorrelation	0.0020
Number of Primary Studies*Fixed Level*Error Variance*Type of Model	0.0019
Series Length*Error Variance*Autocorrelation	0.0018
Series Length*Number of Participants*fix*Error Variance*Type of Model	0.0016
Number of Participants*Number of Primary Studies	0.0016
Number of Primary Studies*Error Variance*Moving Average	0.0014
Series Length*Number of Primary Studies*Error Variance*Moving Average	0.0014
Number of Participants*Moving Average	0.0013
Series Length*Error Variance*Moving Average	0.0013
Error Variance*Moving Average	0.0013
Number of Participants*Number of Primary Studies*Moving Average	0.0013
Series Length*Number of Participants*Number of Primary Studies*Moving Average	0.0013
Series Length*Number of Participants*Moving Average	0.0013
Number of Primary Studies*Fixed Level*Error Variance*Autocorrelation	0.0013
Series Length*Number of Participants*fix*Autocorrelation*Type of Model	0.0013
Number of Participants*Error Variance	0.0012
Series Length*Number of Primary Studies*Autocorrelation	0.0012
Number of Participants*Fixed Level*Error Variance	0.0012
Number of Primary Studies*Fixed Level*Type of Model	0.0011
Series Length*Type of Model	0.0011
Number of Primary Studies*Fixed Level*Autocorrelation*Type of Model	0.0011
Number of Primary Studies*Autocorrelation	0.0011
Number of Participants*Number of Primary Studies*fix*Error Variance*Autocorrelation	0.0010
Number of Participants*Fixed Level*Autocorrelation*Type of Model	0.0010
Series Length*Number of Participants	0.0008
Series Length	0.0008
Number of Participants*Number of Primary Studies*Fixed Level*Type of Model	0.0008
Number of Participants*Number of Primary Studies*Error Variance	0.0008
Number of Participants*Error Variance*Autocorrelation	0.0008
Series Length*Number of Primary Studies*Fixed Level*Autocorrelation	0.0007
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.0007
Series Length*Number of Primary Studies*Moving Average	0.0007
Series Length*Fixed Level*Error Variance*Type of Model	0.0007
Number of Primary Studies*Moving Average	0.0007
Series Length*Number of Participants*Error Variance*Moving Average	0.0006
Number of Participants*Error Variance*Moving Average	0.0006
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Moving Average	0.0006
Series Length*Number of Primary Studies	0.0006
Number of Participants*Number of Primary Studies*Error Variance*Moving Average	0.0006

Table A6 (Continued)

	η^2
Moving Average	0.0006
Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	0.0006
Series Length*Number of Participants*Number of Primary Studies* Fixed Level *Error Variance	0.0006
Series Length*Moving Average	0.0006
Number of Participants*Number of Primary Studies*Fixed Level*Error Variance	0.0006
Series Length*Fixed Level*Autocorrelation	0.0005
Fixed Level*Type of Model	0.0005
Number of Primary Studies*Error Variance*Autocorrelation	0.0005
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.0005
Number of Participants*Fixed Level*Moving Average*Type of Model	0.0004
Fixed Level*Error Variance	0.0004
Series Length*Number of Participants* Fixed Level *the*Type of Model	0.0004
Series Length*Number of Participants*Number of Primary Studies	0.0004
Series Length*Fixed Level*Error Variance*Autocorrelation	0.0004
Series Length*Number of Participants*Fixed Level*Type of Model	0.0004
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation	0.0004
Series Length*Number of Participants*Autocorrelation	0.0004
Number of Primary Studies*fix*Error Variance*the*Type of Model	0.0004
Number of Participants*Number of Primary Studies*Autocorrelation	0.0004
Series Length*fix*Error Variance*the*Type of Model	0.0004
Series Length*Number of Participants*Error Variance*Type of Model	0.0004
Fixed Level*Error Variance*Moving Average*Type of Model	0.0003
Fixed Level*Autocorrelation*Type of Model	0.0003
Series Length*Number of Participants*fix*Error Variance*Autocorrelation	0.0003
Number of Participants*Number of Primary Studies*fix*the*Type of Model	0.0003
Number of Primary Studies*Fixed Level*Error Variance	0.0003
Series Length*Number of Participants*Fixed Level*Error Variance	0.0002
Series Length*Number of Participants*Fixed Level*Autocorrelation	0.0002
Number of Primary Studies*Fixed Level	0.0002
Number of Participants*Fixed Level*Autocorrelation	0.0001
Number of Primary Studies*Fixed Level*Autocorrelation	0.0001
Series Length*Number of Primary Studies*Fixed Level*Type of Model	0.0001
Number of Participants*Number of Primary Studies*Fixed Level	0.0001
Number of Participants*Fixed Level*Type of Model	0.0001
Series Length*Error Variance*Type of Model	0.0001
Series Length*Fixed Level*Error Variance	0.0001
Fixed Level	0.0001
Series Length*Number of Participants*Number of Primary Studies*Error Variance	9.67E-5
Series Length*Number of Participants*Fixed Level*Moving Average	8.41E-5

TableA6 (Continued)

	η2
Number of Participants*Fixed Level*Moving Average	8.33E-5
Series Length*Number of Participants*Fixed Level	7.14E-5
Series Length*Fixed Level*Error Variance*Moving Average	6.73E-5
Number of Primary Studies*Fixed Level*Error Variance*Moving Average	6.69E-5
Series Length*Number of Primary Studies*fix*Error Variance*Moving Average	6.54E-5
Fixed Level*Error Variance*Moving Average	6.35E-5
Series Length*Number of Participants*Number of Primary Studies*fix*Moving Average	6.25E-5
Series Length*Number of Participants*Error Variance	5.99E-5
Fixed Level*Autocorrelation	5.94E-5
Number of Participants*Number of Primary Studies*Fixed Level*Moving Average	5.66E-5
Series Length*Fixed Level*Type of Model	5.52E-5
Number of Participants*fix*Error Variance*the*Type of Model	5E-5
Series Length*Fixed Level*Moving Average*Type of Model	4.72E-5
Fixed Level*Moving Average*Type of Model	4.36E-5
Number of Primary Studies*Fixed Level*Moving Average*Type of Model	3.93E-5
Series Length*Number of Primary Studies*fix*the*Type of Model	3.68E-5
Number of Participants*Fixed Level	2.84E-5
Series Length*Number of Primary Studies*Fixed Level	2.29E-5
Series Length*Number of Primary Studies*Error Variance*Type of Model	1.58E-5
Number of Participants*Number of Primary Studies*fix*Error Variance*Moving Average	1.44E-5
Series Length*Number of Primary Studies*fix*Error Variance*Type of Model	1.32E-5
Series Length*Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	8.91E-6
Fixed Level*Moving Average	8.55E-6
Number of Primary Studies*Fixed Level*Moving Average	8.29E-6
Series Length*Fixed Level	8.24E-6
Series Length*Fixed Level*Moving Average	8.23E-6
Series Length*Number of Primary Studies*Fixed Level*Moving Average	7.41E-6
Series Length*Number of Participants*Number of Primary Studies*fix*Type of Model	6.48E-6
Series Length*Number of Primary Studies*Error Variance	6.11E-6
Series Length*Number of Participants*fix*Error Variance*Moving Average	6.03E-6
Number of Participants*Fixed Level*Error Variance*Moving Average	6E-6
Series Length*Number of Primary Studies*fix*Error Variance*Autocorrelation	5.67E-6
Series Length*Number of Participants*Number of Primary Studies*Fixed Level	1.93E-6
Series Length*Error Variance	8.75E-7
Series Length*Number of Primary Studies*Fixed Level*Error Variance	6.8E-7
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0

Table A6 (Continued)

	η2
Series Length*Number of Primary Studies*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Series Length*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*Error Variance*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Error Variance*Autocorrelation*Moving Average	0
Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Series Length*Fixed Level*Autocorrelation*Moving Average	0
Number of Participants*Fixed Level*Autocorrelation*Moving Average	0
Series Length*Number of Participants*fix*Autocorrelation*Moving Average	0
Number of Primary Studies*Fixed Level*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*fix*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*fix*Autocorrelation*Moving Average	0
Fixed Level*Error Variance*Autocorrelation*Moving Average	0
Series Length*fix*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*fix*Error Variance*Autocorrelation*Moving Average	0
Number of Primary Studies*fix*Error Variance*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Series Length*Autocorrelation*Moving Average*Type of Model	0
Number of Participants*Autocorrelation*Moving Average*Type of Model	0
Series Length*Number of Participants*Autocorrelation*the*Type of Model	0
Number of Primary Studies*Autocorrelation*Moving Average*Type of Model	0
Series Length*Number of Primary Studies*Autocorrelation*the*Type of Model	0
Number of Participants*Number of Primary Studies*Autocorrelation*the*Type of Model	0
Error Variance*Autocorrelation*Moving Average*Type of Model	0
Series Length*Error Variance*Autocorrelation*the*Type of Model	0
Number of Participants*Error Variance*Autocorrelation*the*Type of Model	0
Number of Primary Studies*Error Variance*Autocorrelation*the*Type of Model	0
Fixed Level*Autocorrelation*Moving Average*Type of Model	0
Series Length*fix*Autocorrelation*the*Type of Model	0
Number of Participants*fix*Autocorrelation*the*Type of Model	0
Number of Primary Studies*fix*Autocorrelation*the*Type of Model	0
fix*Error Variance*Autocorrelation*the*Type of Model	0

Table A7

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI Width for the Shift in Slope

	η^2
Autocorrelation*Type of Model	0.11118
Number of Primary Studies	0.09697
Error Variance	0.09252
Type of Model	0.06687
Error Variance*Autocorrelation*Type of Model_ca	0.03199
Error Variance*Type of Model	0.02373
Number of Participants*fix*Error Variance*Autocorrelation*Type of Model	0.01699
Autocorrelation	0.01528
Series Length*Number of Participants*Error Variance*Autocorrelation*Type of Model	0.01513
Number of Participants*Type of Model	0.0132
Fixed Level*Error Variance*Autocorrelation*Type of Model	0.01297
Number of Primary Studies*Error Variance*Moving Average*Type of Model	0.01285
Series Length*Number of Primary Studies*Error Variance*the*Type of Model	0.01282
Number of Participants*Autocorrelation*Type of Model	0.01212
Series Length	0.01211
Number of Participants*Moving Average*Type of Model	0.01193
Series Length*Error Variance*Moving Average*Type of Model	0.01192
Error Variance*Moving Average*Type of Model	0.01189
Series Length*Number of Participants*Moving Average*Type of Model	0.01187
Series Length*Number of Participants*Number of Primary Studies*the*Type of Model	0.01179
Series Length*Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.01179
Number of Participants*Number of Primary Studies*Moving Average*Type of Model	0.01171
Series Length*Type of Model	0.01168
Number of Primary Studies*Type of Model	0.01167
Number of Primary Studies*Error Variance	0.0102
Number of Participants	0.00919
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.0081
Series Length*Number of Primary Studies*Autocorrelation*Type of Model	0.00796
Series Length*Autocorrelation*Type of Model	0.0077
Number of Primary Studies*Autocorrelation*Type of Model	0.00762
Series Length*Number of Participants*Type of Model	0.00755
Number of Participants*Error Variance*Type of Model	0.00731
Number of Participants*Fixed Level*Error Variance*Type of Model	0.0069
Number of Participants*Error Variance*Autocorrelation*Type of Model	0.00679
Series Length*Number of Primary Studies*Moving Average*Type of Model	0.00655
Number of Primary Studies*Moving Average*Type of Model	0.00649
Series Length*Number of Participants*Error Variance*the*Type of Model	0.0063
Number of Participants*Error Variance*Moving Average*Type of Model	0.00627

Table A7 (Continued)

	η^2
Number of Primary Studies*fix*Error Variance*Autocorrelation*Type of Model	0.0062
Number of Participants*Number of Primary Studies*Error Variance*the*Type of Model	0.00618
Series Length*fix*Error Variance*Autocorrelation*Type of Model	0.00599
Moving Average*Type of Model	0.00588
Series Length*Moving Average*Type of Model	0.00584
Series Length*Number of Primary Studies*Type of Model	0.00576
Error Variance*Autocorrelation	0.00563
Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.00537
Number of Participants*Number of Primary Studies*Type of Model	0.00517
Series Length*Number of Participants*Autocorrelation*Type of Model	0.005
Number of Participants*Number of Primary Studies*fix*Error Variance*Type of Model	0.00459
Number of Primary Studies*Error Variance*Type of Model	0.00423
Number of Participants*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.00418
Series Length*Error Variance*Autocorrelation*Type of Model	0.004144
Series Length*Number of Participants*Number of Primary Studies*Type of Model	0.004042
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.003828
Series Length*Error Variance*Type of Model	0.003781
Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.003736
Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.003489
Series Length*Number of Participants*fix*Error Variance*Type of Model	0.003118
Series Length*Number of Participants	0.002905
Number of Participants*Fixed Level*Error Variance*Autocorrelation	0.002869
Fixed Level*Error Variance*Type of Model	0.002852
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.002671
Series Length*Number of Primary Studies	0.002584
Series Length*Number of Participants*Error Variance*Autocorrelation	0.002504
Series Length*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.00232
Fixed Level*Error Variance*Autocorrelation	0.002178
Series Length*Number of Primary Studies*Error Variance*Moving Average	0.002164
Number of Primary Studies*Error Variance*Moving Average	0.002123
Series Length*Number of Participants*Error Variance*Type of Model	0.002067
Series Length*Error Variance*Moving Average	0.002014
Number of Participants*Moving Average	0.001995
Number of Participants*Number of Primary Studies	0.00198
Series Length*Number of Participants*Number of Primary Studies*Moving Average	0.001978
Series Length*Number of Participants*Moving Average	0.001973
Number of Participants*Number of Primary Studies*Moving Average	0.00197
Series Length*Number of Primary Studies*Error Variance*Autocorrelation	0.001968
Error Variance*Moving Average	0.001964

Table A7 (Continued)

	η^2
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.001915
Series Length*Fixed Level*Autocorrelation*Type of Model	0.001864
Number of Participants*Autocorrelation	0.001695
Number of Participants*Number of Primary Studies*fix*Error Variance*Autocorrelation	0.00148
Number of Participants*Error Variance	0.00138
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.001351
Series Length*Number of Primary Studies*Autocorrelation	0.001299
Number of Primary Studies*Fixed Level*Error Variance*Type of Model	0.00127
Series Length*Fixed Level*Error Variance*Type of Model	0.001264
Number of Participants*Error Variance*Autocorrelation	0.001153
Number of Participants*Fixed Level*Error Variance	0.00113
Series Length*Number of Primary Studies*Moving Average	0.001126
Number of Primary Studies*Moving Average	0.001105
Series Length*Number of Participants*Error Variance*Moving Average	0.001053
Number of Participants*Error Variance*Moving Average	0.001052
Number of Primary Studies*Fixed Level*Error Variance*Autocorrelation	0.001041
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Moving Average	0.001035
Number of Participants*Number of Primary Studies*Error Variance*Moving Average	0.001035
Number of Primary Studies*Autocorrelation	0.001033
Series Length*Moving Average	0.001014
Series Length*Fixed Level*Error Variance*Autocorrelation	0.000991
Moving Average	0.000986
Series Length*Number of Participants*Number of Primary Studies	0.000984
Series Length*Autocorrelation	0.000965
Series Length*Number of Participants*fix*Error Variance*Autocorrelation	0.000956
Number of Participants*Number of Primary Studies*Error Variance	0.000949
Number of Participants*Fixed Level*Autocorrelation*Type of Model	0.000811
Number of Participants*Number of Primary Studies*Fixed Level*Type of Model	0.0008
Series Length*Number of Participants*Number of Primary Studies*fix*Error Variance	0.000766
Series Length*Number of Participants*Autocorrelation	0.000751
Number of Participants*Number of Primary Studies*Fixed Level*Error Variance	0.000746
Series Length*Error Variance*Autocorrelation	0.000718
Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	0.000695
Number of Primary Studies*Fixed Level*Type of Model_	0.000653
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation	0.000641
Number of Primary Studies*Error Variance*Autocorrelation	0.000599
Number of Participants*Number of Primary Studies*Autocorrelation	0.000579
Series Length*Number of Participants*Fixed Level*Error Variance	0.000512
Fixed Level*Error Variance	0.000484

Table A7 (Continued)

	η^2
Number of Primary Studies*Fixed Level*Autocorrelation*Type of Model	0.000446
Number of Participants*Fixed Level*Moving Average*Type of Model	0.000445
Series Length*Number of Primary Studies*Error Variance*Type of Model	0.00044
Series Length*Number of Participants*fix*the*Type of Model	0.000439
Series Length*Number of Primary Studies*Fixed Level*Autocorrelation	0.000384
Number of Primary Studies*fix*Error Variance*the*Type of Model	0.000364
Fixed Level*Type of Model	0.000355
Series Length*Number of Participants*fix*Autocorrelation*Type of Model	0.000353
Series Length*fix*Error Variance*the*Type of Model	0.000338
Fixed Level*Autocorrelation*Type of Mode	0.000336
Fixed Level*Error Variance*Moving Average*Type of Model	0.000327
Series Length*Fixed Level*Autocorrelation	0.000313
Number of Participants*Number of Primary Studies*fix*the*Type of Model	0.00028
Series Length*Number of Participants*Number of Primary Studies*Error Variance	0.000275
Series Length*Number of Participants*Error Variance	0.000262
Series Length*Number of Participants*Fixed Level*Type of Model	0.000258
Series Length*Error Variance	0.000216
Series Length*Fixed Level*Error Variance	0.000211
Number of Primary Studies*Fixed Level*Error Variance	0.000209
Number of Participants*Number of Primary Studies*Fixed Level	0.000141
Number of Participants*fix*Error Variance*the*Type of Model	0.000138
Number of Participants*Fixed Level*Autocorrelation	0.000137
Series Length*Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	0.000136
Series Length*Fixed Level*Moving Average*Type of Model	0.000135
Fixed Level*Moving Average*Type of Model	0.000132
Number of Primary Studies*Fixed Level*Moving Average*Type of Model	0.000119
Series Length*Number of Primary Studies*fix*the*Type of Model	0.000117
Series Length*Number of Participants*Number of Primary Studies*fix*Type of Model	0.000106
Number of Primary Studies*Fixed Level	0.0001
Number of Participants*Fixed Level*Moving Average	7.71E-05
Number of Primary Studies*Fixed Level*Autocorrelation	7.42E-05
Series Length*Number of Participants*Fixed Level*Moving Average	7.09E-05
Number of Primary Studies*Fixed Level*Error Variance*Moving Average	6.21E-05
Series Length*Number of Participants*Fixed Level*Autocorrelation	6.13E-05
Series Length*Number of Primary Studies*fix*Error Variance*Autocorrelation	6.05E-05
Series Length*Number of Primary Studies*fix*Error Variance*Moving Average	5.85E-05
Series Length*Fixed Level*Error Variance*Moving Average	5.73E-05
Fixed Level	5.69E-05
Fixed Level*Error Variance*Moving Average	5.64E-05
Fixed Level*Autocorrelation	5.38E-05

Table A7 (Continued)

	η^2
Number of Participants*Number of Primary Studies*Fixed Level*Moving Average	4.93E-05
Series Length*Number of Participants*Number of Primary Studies*fix*Moving Average	4.6E-05
Number of Participants*Fixed Level*Type of Model	4.44E-05
Number of Participants*Number of Primary Studies*fix*Error Variance*Moving Average	4.17E-05
Series Length*Number of Participants*Fixed Level	4.08E-05
Series Length*Number of Primary Studies*Error Variance	3.4E-05
Series Length*Number of Primary Studies*fix*Error Variance*Type of Model	2.84E-05
Number of Participants*Fixed Level*Error Variance*Moving Average	2.23E-05
Series Length*Fixed Level*Moving Average	2.07E-05
Series Length*Number of Participants*fix*Error Variance*Moving Average	2.05E-05
Fixed Level*Moving Average	2.05E-05
Series Length*Number of Primary Studies*Fixed Level*Moving Average	1.9E-05
Series Length*Number of Participants*Number of Primary Studies*Fixed Level	1.9E-05
Number of Primary Studies*Fixed Level*Moving Average	1.85E-05
Number of Participants*Fixed Level	9.33E-06
Series Length*Number of Primary Studies*Fixed Level*Type of Model	7.77E-06
Series Length*Number of Primary Studies*Fixed Level*Error Variance	5.42E-06
Series Length*Number of Primary Studies*Fixed Level	1.35E-06
Series Length*Fixed Level*Type of Model	4.18E-07
Series Length*Fixed Level	6.1E-08
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Series Length*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*Error Variance*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Error Variance*Autocorrelation*Moving Average	0
Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Series Length*Fixed Level*Autocorrelation*Moving Average	0
Number of Participants*Fixed Level*Autocorrelation*Moving Average	0
Series Length*Number of Participants*fix*Autocorrelation*Moving Average	0

Table A7 (Continued)

	η2
Number of Primary Studies*Fixed Level*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*fix*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*fix*Autocorrelation*Moving Average	0
Fixed Level*Error Variance*Autocorrelation*Moving Average	0
Series Length*fix*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*fix*Error Variance*Autocorrelation*Moving Average	0
Number of Primary Studies*fix*Error Variance*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Series Length*Autocorrelation*Moving Average*Type of Model	0
Number of Participants*Autocorrelation*Moving Average*Type of Model	0
Series Length*Number of Participants*Autocorrelation*the*Type of Model	0
Number of Primary Studies*Autocorrelation*Moving Average*Type of Model	0
Series Length*Number of Primary Studies*Autocorrelation*the*Type of Model	0
Number of Participants*Number of Primary Studies*Autocorrelation*the*Type of Model	0
Error Variance*Autocorrelation*Moving Average*Type of Model	0
Series Length*Error Variance*Autocorrelation*the*Type of Model	0
Number of Participants*Error Variance*Autocorrelation*the*Type of Model	0
Number of Primary Studies*Error Variance*Autocorrelation*the*Type of Model	0
Fixed Level*Autocorrelation*Moving Average*Type of Model	0
Series Length*fix*Autocorrelation*the*Type of Model	0
Number of Participants*fix*Autocorrelation*the*Type of Model	0
Number of Primary Studies*fix*Autocorrelation*the*Type of Model	0
fix*Error Variance*Autocorrelation*the*Type of Model	0
Total Explained	0.9457

Table A8

Eta-Squared Values (η^2)for the Association of the Design Factors with the Power Estimates for the Shift in Level

	η^2
Error Variance	0.30027
Number of Primary Studies	0.2993
Number of Primary Studies*Error Variance	0.29828
Number of Participants	0.02225
Number of Participants*Number of Primary Studies	0.02225
Number of Participants*Error Variance	0.02192
Series Length*Number of Primary Studies	0.00184
Series Length	0.00179
Series Length*Error Variance	0.00174
Number of Primary Studies*Autocorrelation	0.0006
Error Variance*Autocorrelation	0.00045

Table A8 (Continued)

	η^2
Autocorrelation	0.00044
Series Length*Number of Participants	0.00039
Autocorrelation*Type of Model	0.0002
Number of Participants*Moving Average	0.00017
Type of Model	0.00015
Series Length*Moving Average	0.00013
Error Variance*Type of Model	0.00012
Series Length*Autocorrelation	0.00011
Number of Participants*Autocorrelation	0.00006
Number of Primary Studies*Type of Model	0.00005
Number of Participants*Type of Model	0.00002
Moving Average	0.00002
Number of Primary Studies*Moving Average	0.00002
Error Variance*Moving Average	0.00001
Series Length*Type of Model	0.00001
Moving Average*Type of Model	0
Autocorrelation*Moving Average	0
Total Explained	0.9726

Table A9

Eta-Squared Values (η^2)for the Association of the Design Factors with the Power Estimates for the Shift in Slope

	η^2
Number of Primary Studies	0.51153
Error Variance	0.41569
Series Length	0.03141
Number of Participants	0.01529
Autocorrelation	0.00129
Type of Model	0.00002
Moving Average	0.00001
Total Explained	0.9752

Table A10

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Level-three Phase Effect

	η^2
Error Variance	0.096415
Number of Primary Studies	0.093087
Type of Model	0.091423
Error Variance*Type of Model	0.04376
Number of Participants	0.042611
Number of Participants*Number of Primary Studies	0.041102
Number of Primary Studies*Error Variance	0.034378
Autocorrelation*Type of Model	0.031906
Number of Participants*Error Variance	0.02619
Number of Primary Studies*Fixed Level*Error Variance*Moving Average	0.018903
Series Length	0.018586
Series Length*Number of Participants*Number of Primary Studies	0.016809
Error Variance*Autocorrelation*Type of Model	0.01632
Autocorrelation	0.016299
Number of Primary Studies*Fixed Level*Autocorrelation	0.016202
Series Length*Number of Participants*Autocorrelation	0.015133
Error Variance*Autocorrelation	0.014939
Number of Participants*Number of Primary Studies*Moving Average	0.013593
Series Length*Number of Primary Studies*Fixed Level*Error Variance	0.012667
Number of Participants*Number of Primary Studies*Fixed Level	0.012435
Number of Primary Studies*Error Variance*Autocorrelation	0.011813
Number of Participants*Moving Average	0.010852
Series Length*Type of Model	0.010326
Number of Primary Studies*Fixed Level	0.010182
Series Length*Error Variance	0.009611
Number of Participants*Fixed Level*Error Variance*Moving Average	0.009564
Number of Participants*Number of Primary Studies*Autocorrelation	0.009294
Number of Primary Studies*Fixed Level*Moving Average	0.009152
Number of Participants*Fixed Level*Autocorrelation	0.008815
Fixed Level	0.007961
Number of Participants*Number of Primary Studies*Fixed Level*Moving Average	0.007297
Series Length*Number of Primary Studies	0.00627
Number of Primary Studies*Fixed Level*Error Variance	0.00603
Number of Participants*Number of Primary Studies*Error Variance*Moving Average	0.005729
Series Length*Number of Primary Studies*Error Variance*Moving Average	0.005603
Series Length*Number of Primary Studies*Fixed Level	0.00558
Number of Participants*Type of Model	0.005553

Table A10 (Continued)

	η^2
Series Length*Error Variance*Type of Model	0.005291
Series Length*Number of Primary Studies*Fixed Level*Moving Average	0.005287
Fixed Level*Error Variance	0.005131
Series Length*Number of Participants*Number of Primary Studies*Fixed Level	0.004597
Series Length*Number of Participants	0.004596
Fixed Level*Error Variance*Moving Average	0.0042
Series Length*Number of Primary Studies*Error Variance	0.004134
Series Length*Error Variance*Moving Average	0.004083
Number of Participants*Number of Primary Studies*Error Variance	0.004025
Series Length*Fixed Level*Autocorrelation	0.003922
Number of Participants*Error Variance*Autocorrelation	0.0039
Number of Participants*Autocorrelation*Type of Model	0.003842
Number of Participants*Error Variance*Type of Model	0.003828
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation	0.003815
Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	0.003713
Series Length*Fixed Level*Error Variance*Autocorrelation	0.003572
Series Length*Autocorrelation*Type of Model	0.003385
Number of Primary Studies*Fixed Level*Error Variance*Autocorrelation	0.003128
Series Length*Error Variance*Autocorrelation	0.003094
Series Length*Number of Participants*Fixed Level	0.003053
Series Length*Number of Primary Studies*Autocorrelation	0.002959
Series Length*Autocorrelation	0.002831
Series Length*Number of Participants*Error Variance*Autocorrelation	0.002515
Series Length*Moving Average	0.002417
Number of Participants*Number of Primary Studies*Fixed Level*Error Variance	0.002409
Series Length*Number of Primary Studies*Error Variance*Autocorrelation	0.002364
Number of Participants*Fixed Level*Error Variance	0.002323
Fixed Level*Autocorrelation	0.002187
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.002022
Number of Primary Studies*Autocorrelation	0.001915
Moving Average	0.001882
Number of Primary Studies*Moving Average	0.00176
Series Length*Number of Primary Studies*Fixed Level*Autocorrelation	0.001723
Number of Participants*Autocorrelation	0.001558
Number of Primary Studies*Error Variance*Moving Average	0.001348
Fixed Level*Moving Average	0.001296
Series Length*Number of Participants*Number of Primary Studies*Moving Average	0.001221
Series Length*Error Variance*Autocorrelation*fixed	0.00116
Number of Participants*Error Variance*Autocorrelation*fixed	0.001109
Number of Participants*Error Variance*Moving Average	0.001084
Series Length*Number of Participants*Type of Model	0.001073

Table A10 (Continued)

	η^2
Error Variance*Moving Average	0.001061
Number of Participants*Fixed Level*Moving Average	0.001009
Fixed Level*Error Variance*Autocorrelation*fixed	0.000917
Fixed Level*Error Variance*Autocorrelation	0.000902
Series Length*Number of Participants*Fixed Level*Moving Average	0.000829
Number of Primary Studies*Error Variance*Type of Model	0.000811
Series Length*Number of Participants*Error Variance*Moving Average	0.000753
Number of Participants*Number of Primary Studies*Autocorrelation*fixed	0.000705
Series Length*Fixed Level*Error Variance*Moving Average	0.000686
Series Length*Fixed Level*Moving Average	0.000674
Series Length*Number of Participants*Moving Average	0.000651
Number of Participants*Fixed Level	0.000629
Number of Participants*Fixed Level*Error Variance*Autocorrelation	0.000556
Number of Participants*Number of Primary Studies*Type of Model	0.000511
Series Length*Number of Primary Studies*Moving Average	0.000479
Series Length*Number of Participants*Fixed Level*Error Variance	0.000475
Number of Participants*Fixed Level*Autocorrelation*fixed	0.000412
Series Length*Number of Participants*Number of Primary Studies*Error Variance	0.000389
Series Length*Number of Primary Studies*Autocorrelation*fixed	0.000351
Number of Primary Studies*Autocorrelation*Type of Model	0.000339
Number of Primary Studies*Fixed Level*Type of Model	0.000328
Number of Primary Studies*Error Variance*Moving Average*fixed	0.000315
Number of Primary Studies*Type of Model	0.000303
Number of Primary Studies*Moving Average*Type of Model	0.000256
Total Explained	0.9465

Table A11

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Level-three Interaction Effect

	η^2
Type of Model	0.14873
Series Length	0.11207
Series Length*Type of Model	0.06868
Error Variance	0.05991
Autocorrelation*Type of Model	0.05601
Error Variance*Type of Model	0.04917
Series Length*Error Variance	0.04066
Number of Participants	0.03706
Number of Primary Studies	0.0361
Series Length*Number of Primary Studies	0.03274

Table A11 (Continued)

	η^2
Series Length*Autocorrelation*Type of Model	0.03112
Error Variance*Autocorrelation*Type of Model	0.02632
Number of Participants*Number of Primary Studies	0.02151
Number of Participants*Error Variance	0.02102
Series Length*Number of Participants	0.02091
Series Length*Error Variance*Type of Model	0.02022
Autocorrelation	0.01911
Series Length*Number of Primary Studies*Error Variance	0.01439
Number of Primary Studies*Error Variance	0.01277
Series Length*Number of Participants*Error Variance	0.01165
Series Length*Number of Participants*Number of Primary Studies	0.01122
Number of Participants*Number of Primary Studies*Error Variance	0.01001
Series Length*Autocorrelation	0.00375
Number of Primary Studies*Autocorrelation	0.00365
Series Length*Number of Participants*Autocorrelation	0.0034
Error Variance*Autocorrelation	0.00339
Number of Participants*Fixed Level	0.0033
Series Length*Number of Primary Studies*Autocorrelation	0.00322
Number of Primary Studies*Moving Average	0.00311
Series Length*Moving Average	0.00288
Number of Primary Studies*Error Variance*Autocorrelation	0.00286
Series Length*Error Variance*Autocorrelation	0.00276
Number of Participants*Number of Primary Studies*Autocorrelation	0.00237
Number of Participants*Number of Primary Studies*Moving Average	0.00233
Series Length*Error Variance*Moving Average	0.00232
Number of Participants*Type of Model	0.00212
Number of Participants*Number of Primary Studies*Fixed Level	0.00211
Number of Primary Studies*Fixed Level	0.00206
Series Length*Number of Primary Studies*Type of Model	0.00198
Series Length*Number of Primary Studies*Moving Average	0.00187
Number of Primary Studies*Type of Model	0.00173
Number of Participants*Number of Primary Studies*Type of Model	0.00172
Fixed Level*Error Variance*Autocorrelation	0.00169
Number of Participants*Autocorrelation*Type of Model	0.00157
Number of Primary Studies*Error Variance*Moving Average	0.00155
Number of Primary Studies*Error Variance*Type of Model	0.00137
Number of Participants*Fixed Level*Autocorrelation	0.00132
Series Length*Fixed Level*Autocorrelation	0.00129
Number of Participants*Fixed Level*Error Variance	0.00124
Fixed Level*Autocorrelation	0.00122
Number of Primary Studies*Autocorrelation*Type of Model	0.00121

Table A11 (Continued)

	η^2
Series Length*Fixed Level*Moving Average	0.001182
Fixed Level*Moving Average	0.00116
Fixed Level*Error Variance*Moving Average	0.000989
Series Length*Number of Primary Studies*Fixed Level	0.000957
Error Variance*Moving Average	0.000894
Number of Primary Studies*Fixed Level*Autocorrelation	0.000706
Number of Participants*Error Variance*Moving Average	0.000682
Series Length*Number of Participants*Fixed Level	0.000608
Series Length*Number of Participants*Moving Average	0.000412
Series Length*Fixed Level*Error Variance	0.000369
Series Length*Number of Participants*Type of Model	0.000364
Number of Participants*Autocorrelation	0.000302
Number of Participants*Error Variance*Autocorrelation	0.000247
Number of Participants*Moving Average	0.000199
Number of Participants*Error Variance*Type of Model	0.000195
Number of Participants*Fixed Level*Moving Average	0.000183
Fixed Level*Moving Average*Type of Model	0.000119
Fixed Level*Error Variance	0.000118
Number of Primary Studies*Fixed Level*Moving Average	0.000108
Number of Primary Studies*Fixed Level*Type of Model	9.86E-05
Fixed Level	7.97E-05
Fixed Level*Type of Model	6.24E-05
Fixed Level*Autocorrelation*Type of Model	5.91E-05
Series Length*Moving Average*Type of Model	5.08E-05
Fixed Level*Error Variance*Type of Model	5.07E-05
Number of Participants*Moving Average*Type of Model	4.11E-05
Error Variance*Moving Average*Type of Model	2.94E-05
Number of Primary Studies*Moving Average*Type of Model	2.37E-05
Series Length*Fixed Level*Type of Model	2.16E-05
Number of Primary Studies*Fixed Level*Error Variance	1.99E-05
Number of Participants*Fixed Level*Type of Model	1.86E-05
Moving Average*Type of Model	1.6E-05
Moving Average	6.91E-06
Series Length*Fixed Level	5.14E-06
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0

Total Explained	0.9371

Table A12

Eta-Squared Values (η^2)for the Association of the Design Factors with the Bias Estimates for the Level-two Phase Effect

	η^2
Type of Model	0.55782
Autocorrelation*Type of Model	0.22198
Error Variance*Autocorrelation	0.04332
Error Variance*Type of Model	0.04332
Error Variance*Autocorrelation*Type of Model	0.0431
Autocorrelation	0.04129
Series Length*Error Variance	0.00251
Number of Primary Studies*Fixed Level*Moving Average	0.00238
Series Length*Number of Primary Studies	0.00191
Number of Participants*Error Variance	0.00184
Fixed Level*Error Variance	0.00151
Number of Participants*Fixed Level*Autocorrelation	0.00142
Fixed Level	0.00116
Series Length*Fixed Level*Moving Average	0.0011
Series Length*Number of Participants	0.00107
Number of Participants*Fixed Level*Error Variance	0.00106
Fixed Level*Error Variance*Autocorrelation	0.00099
Fixed Level*Autocorrelation	0.00085
Error Variance*Moving Average	0.00082
Number of Participants*Fixed Level	0.00082
Series Length*Autocorrelation*Type of Model	0.00076
Number of Participants*Number of Primary Studies	0.00065
Series Length*Error Variance*Autocorrelation	0.00063
Number of Primary Studies*Error Variance	0.00062
Fixed Level*Error Variance*Moving Average	0.00057
Series Length*Number of Participants*Moving Average	0.00054
Number of Participants*Number of Primary Studies*Moving Average	0.00052
Number of Participants*Moving Average*Type of Model	0.0005
Series Length*Fixed Level	0.00049
Series Length*Number of Primary Studies*Autocorrelation	0.00048
Number of Primary Studies*Moving Average	0.00048
Number of Primary Studies*Autocorrelation	0.00046
Fixed Level*Moving Average*Type of Model	0.00043
Moving Average*Type of Model	0.00039
Number of Participants*Error Variance*Autocorrelation	0.00039
Number of Primary Studies*Autocorrelation*Type of Model	0.00039

Table A12 (Continued)

	η^2
Series Length*Number of Participants*Type of Model	0.00037
Number of Participants*Autocorrelation	0.00036
Number of Participants*Number of Primary Studies*Autocorrelation	0.00035
Number of Participants*Fixed Level*Moving Average	0.00035
Number of Participants*Autocorrelation*Type of Model	0.00034
Number of Primary Studies*Error Variance*Type of Model	0.00033
Number of Participants*Error Variance*Type of Model	0.00031
Series Length*Fixed Level*Autocorrelation	0.00029
Number of Primary Studies*Error Variance*Autocorrelation	0.00027
Series Length*Type of Model	0.00027
Number of Participants*Number of Primary Studies*Type of Model	0.00027
Moving Average	0.00025
Number of Participants*Error Variance*Moving Average	0.00024
Series Length*Number of Participants*Fixed Level	0.00023
Number of Primary Studies*Moving Average*Type of Model	0.00023
Number of Primary Studies*Type of Model	0.000217
Fixed Level*Moving Average	0.000217
Number of Primary Studies*Error Variance*Moving Average	0.000216
Number of Participants*Number of Primary Studies*Fixed Level	0.000216
Fixed Level*Error Variance*Type of Model	0.000185
Series Length*Autocorrelation	0.000182
Fixed Level*Autocorrelation*Type of Model	0.000173
Error Variance*Moving Average*Type of Model	0.00017
Series Length*Moving Average*Type of Model	0.000167
Number of Participants*Moving Average	0.000165
Fixed Level*Type of Model	0.000157
Series Length	0.000148
Series Length*Number of Participants*Error Variance	0.000139
Number of Participants*Fixed Level*Type of Model	0.000127
Number of Participants*Number of Primary Studies*Error Variance	0.000119
Series Length*Number of Primary Studies*Moving Average	0.000103
Series Length*Fixed Level*Error Variance	0.000101
Number of Participants*Type of Model	9.7E-05
Series Length*Error Variance*Type of Model	9.67E-05
Series Length*Number of Primary Studies*Type of Model	8.87E-05
Number of Primary Studies*Fixed Level*Type of Model	7.72E-05
Series Length*Number of Participants*Autocorrelation	6.92E-05
Number of Primary Studies	6.42E-05
Series Length*Fixed Level*Type of Model	5.55E-05
Series Length*Number of Participants*Number of Primary Studies	4.84E-05
Series Length*Moving Average	4.66E-05

Table A12 (Continued)

	η^2
Number of Primary Studies*Fixed Level*Autocorrelation	3.32E-05
Number of Primary Studies*Fixed Level*Error Variance	3.05E-05
Series Length*Error Variance*Moving Average	2.14E-05
Error Variance	9.18E-06
Number of Primary Studies*Fixed Level	2.93E-06
Series Length*Number of Primary Studies*Fixed Level	1.5E-06
Number of Participants	9.9E-07
Series Length*Number of Primary Studies*Error Variance	5.59E-07
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9866

Table A13

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Level-two Interaction Effect

	η^2
Type of Model	0.44809
Autocorrelation*Type of Model	0.19139
Series Length*Error Variance*Autocorrelation	0.06297
Series Length*Error Variance*Type of Model	0.05562
Autocorrelation	0.0304
Series Length*Autocorrelation*Type of Model	0.02342
Series Length*Type of Model	0.01366
Series Length*Number of Participants*Error Variance	0.01311
Series Length*Number of Primary Studies*Error Variance	0.0127
Series Length*Number of Primary Studies	0.01141
Error Variance*Autocorrelation*Type of Model	0.01079
Series Length*Number of Participants	0.00956
Number of Primary Studies*Error Variance	0.00908
Number of Participants*Error Variance	0.00838
Series Length*Number of Participants*Number of Primary Studies	0.00749
Series Length*Error Variance	0.00692
Error Variance*Type of Model	0.0065
Series Length*Number of Primary Studies*Autocorrelation	0.00506
Number of Primary Studies*Autocorrelation	0.00489

Table A13 (Continued)

	η^2
Number of Primary Studies	0.00334
Number of Primary Studies*Type of Model	0.00306
Number of Participants*Number of Primary Studies	0.00306
Error Variance*Autocorrelation	0.00303
Number of Participants*Number of Primary Studies*Error Variance	0.00269
Number of Participants	0.00258
Series Length*Autocorrelation	0.00246
Fixed Level*Autocorrelation	0.00201
Number of Primary Studies*Error Variance*Autocorrelation	0.00201
Number of Participants*Error Variance*Autocorrelation	0.00147
Number of Primary Studies*Autocorrelation*Type of Model	0.00143
Number of Participants*Fixed Level*Error Variance	0.00139
Number of Participants*Type of Model	0.00129
Series Length*Number of Primary Studies*Type of Model	0.00125
Series Length*Number of Participants*Autocorrelation	0.00118
Number of Participants*Number of Primary Studies*Autocorrelation	0.0011
Number of Participants*Autocorrelation	0.0011
Error Variance	0.00109
Fixed Level	0.00095
Moving Average	0.00088
Fixed Level*Moving Average	0.00076
Series Length*Fixed Level*Moving Average	0.00069
Series Length*Number of Participants*Type of Model	0.00066
Number of Participants*Autocorrelation*Type of Model	0.00058
Number of Participants*Number of Primary Studies*Type of Model	0.00046
Series Length*Number of Primary Studies*Moving Average	0.00044
Series Length	0.00042
Number of Primary Studies*Fixed Level*Autocorrelation	0.00039
Series Length*Fixed Level*Autocorrelation	0.00036
Series Length*Moving Average	0.00036
Series Length*Number of Participants*Moving Average	0.00025
Number of Primary Studies*Fixed Level*Moving Average	0.00025
Number of Primary Studies*Error Variance*Moving Average	0.000235
Number of Primary Studies*Fixed Level	0.000231
Number of Participants*Fixed Level*Autocorrelation	0.0002
Number of Primary Studies*Error Variance*Type of Model	0.000197
Number of Primary Studies*Moving Average	0.000192
Number of Participants*Moving Average	0.000178
Number of Participants*Number of Primary Studies*Fixed Level	0.000167
Fixed Level*Error Variance*Moving Average	0.000166
Fixed Level*Error Variance*Autocorrelation	0.000162

Table A13 (Continued)

	η^2
Series Length*Fixed Level*Error Variance	0.000113
Number of Participants*Number of Primary Studies*Moving Average	9.52E-05
Number of Primary Studies*Fixed Level*Error Variance	8.31E-05
Series Length*Error Variance*Moving Average	7.62E-05
Number of Participants*Fixed Level*Moving Average	6.69E-05
Number of Participants*Error Variance*Type of Model	6.2E-05
Series Length*Number of Participants*Fixed Level	5.39E-05
Number of Participants*Fixed Level	4.88E-05
Number of Participants*Error Variance*Moving Average	4.27E-05
Error Variance*Moving Average	2.77E-05
Fixed Level*Error Variance*Type of Model	2.11E-05
Number of Participants*Fixed Level*Type of Model	2.02E-05
Fixed Level*Moving Average*Type of Model	1.8E-05
Fixed Level*Autocorrelation*Type of Model	1.67E-05
Number of Primary Studies*Moving Average*Type of Model	1.53E-05
Moving Average*Type of Model	1.33E-05
Series Length*Fixed Level*Type of Model	1.04E-05
Number of Primary Studies*Fixed Level*Type of Model	1.04E-05
Number of Participants*Moving Average*Type of Model	1.01E-05
Error Variance*Moving Average*Type of Model	8.5E-06
Series Length*Moving Average*Type of Model	8.07E-06
Fixed Level*Type of Model	6.65E-06
Fixed Level*Error Variance	1.67E-06
Series Length*Fixed Level	1.25E-07
Series Length*Number of Primary Studies*Fixed Level	1.1E-07
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9770

Table A14

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Autocorrelation Parameter

	η^2
Autocorrelation*Type of Model	0.24284
Error Variance*Autocorrelation*Type of Model	0.12234
Error Variance*Type of Model	0.08035
Type of Model	0.07695
Error Variance*Autocorrelation	0.06308
Autocorrelation	0.06018
Error Variance	0.05549
Series Length	0.03203
Series Length*Autocorrelation*Type of Model	0.03044
Series Length*Type of Model	0.02916
Series Length*Error Variance*Autocorrelation	0.01736
Series Length*Autocorrelation	0.01645
Number of Primary Studies*Autocorrelation*Type of Model	0.01219
Number of Primary Studies*Autocorrelation	0.01033
Number of Primary Studies*Error Variance*Type of Model	0.00951
Number of Participants*Error Variance*Type of Model	0.00899
Number of Primary Studies*Error Variance	0.00782
Series Length*Error Variance*Type of Model	0.00753
Series Length*Number of Primary Studies	0.00679
Number of Participants*Error Variance	0.00605
Series Length*Number of Primary Studies*Type of Model	0.00393
Number of Primary Studies*Type of Model	0.00364
Number of Participants*Autocorrelation*Type of Model	0.00346
Series Length*Number of Primary Studies*Autocorrelation	0.00341
Series Length*Number of Participants	0.00338
Number of Participants*Autocorrelation	0.00333
Number of Primary Studies*Error Variance*Autocorrelation	0.0029
Number of Participants*Error Variance*Autocorrelation	0.00276
Series Length*Number of Participants*Autocorrelation	0.00191
Series Length*Number of Participants*Number of Primary Studies	0.0018
Series Length*Number of Participants*Type of Model	0.00159
Series Length*Error Variance	0.00151
Number of Participants*Number of Primary Studies*Autocorrelation	0.0013
Number of Participants*Type of Model	0.00129
Number of Participants*Number of Primary Studies	0.00094
Number of Participants*Number of Primary Studies*Type of Model	0.00078
Number of Participants	0.00063

Table A14 (Continued)

	η^2
Number of Primary Studies	0.00042
Number of Participants*Fixed Level*Autocorrelation	0.00034
Fixed Level*Autocorrelation*Type of Model	0.00027
Number of Participants*Number of Primary Studies*Error Variance	0.00022
Number of Participants*Error Variance*Moving Average	0.00017
Fixed Level*Autocorrelation	0.00017
Fixed Level*Error Variance*Moving Average	0.00015
Series Length*Number of Participants*Error Variance	0.00012
Series Length*Number of Participants*Fixed Level	0.00012
Series Length*Fixed Level*Type of Model	0.00011
Fixed Level*Moving Average	0.00011
Series Length*Fixed Level*Moving Average	0.00008
Series Length*Fixed Level*Error Variance	0.00008
Number of Participants*Number of Primary Studies*Moving Average	0.00007
Number of Participants*Moving Average	6.23E-05
Fixed Level*Error Variance*Autocorrelation	4.63E-05
Moving Average	4.59E-05
Series Length*Error Variance*Moving Average	4.5E-05
Number of Primary Studies*Fixed Level*Error Variance	4.45E-05
Number of Primary Studies*Fixed Level*Autocorrelation	4.33E-05
Number of Primary Studies*Fixed Level*Moving Average	3.96E-05
Error Variance*Moving Average	3.29E-05
Number of Primary Studies*Moving Average	2.89E-05
Number of Participants*Moving Average*Type of Model	2.54E-05
Number of Participants*Fixed Level*Moving Average	2.44E-05
Fixed Level*Moving Average*Type of Model	2.31E-05
Number of Participants*Number of Primary Studies*Fixed Level	2.22E-05
Number of Primary Studies*Moving Average*Type of Model	2.15E-05
Series Length*Fixed Level	1.74E-05
Series Length*Number of Primary Studies*Moving Average	1.71E-05
Series Length*Number of Participants*Moving Average	1.64E-05
Moving Average*Type of Model	1.47E-05
Number of Participants*Fixed Level*Error Variance	1.28E-05
Series Length*Moving Average	1.13E-05
Series Length*Moving Average*Type of Model	1.09E-05
Fixed Level*Error Variance*Type of Model	9.86E-06
Fixed Level	6.63E-06
Error Variance*Moving Average*Type of Model	5.6E-06
Fixed Level*Type of Model	5.09E-06
Series Length*Number of Primary Studies*Fixed Level	4.76E-06

Table A14 (Continued)

	η^2
Number of Primary Studies*Fixed Level*Type of Model	4.46E-06
Series Length*Fixed Level*Autocorrelation	2.9E-06
Number of Primary Studies*Fixed Level	2.63E-06
Number of Participants*Fixed Level	2.13E-06
Number of Participants*Fixed Level*Type of Model	2.08E-06
Number of Primary Studies*Error Variance*Moving Average	1.83E-06
Fixed Level*Error Variance	4.26E-07
Series Length*Number of Primary Studies*Error Variance	2.16E-07
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9375

Table A15

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Autocorrelation Parameter

	η^2
Moving Average	0.60692
Type of Model	0.27378
Autocorrelation	0.05027
Error Variance	0.00217
Number of Primary Studies	0.00176
Series Length	0.00071
Number of Participants	0.00049
Fixed Level	0
Total Explained	0.9361

Table A16

Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Level-three Phase Effect

	η^2
Error Variance	0.70311
Number of Primary Studies	0.2014
Number of Participants	0.02262
Series Length	0.00304
Autocorrelation	0.00018
Type of Model	0.00003
Moving Average	0.00002
Fixed Level	0
Total Explained	0.9304

Table A17

Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Level-three Interaction Effect

	η^2
Error Variance	0.66677
Number of Primary Studies	0.1979
Number of Primary Studies*Error Variance	0.05293
Series Length	0.03996
Number of Participants	0.02858
Number of Participants*Error Variance	0.00361
Series Length*Number of Participants	0.00326
Autocorrelation	0.00157
Number of Participants*Number of Primary Studies	0.00113
Series Length*Error Variance	0.001
Series Length*Number of Primary Studies	0.00071
Series Length*Autocorrelation	0.0003
Number of Participants*Autocorrelation	0.00029
Number of Primary Studies*Autocorrelation	0.00011
Error Variance*Autocorrelation	0.00006
Autocorrelation*Type of Model	0.00003
Type of Model	0.00003
Series Length*Fixed Level	0.00002
Series Length*Type of Model	0.00002
Fixed Level*Error Variance	0.00001
Number of Participants*Moving Average	0.00001
Series Length*Moving Average	0.00001
Fixed Level*Autocorrelation	0
Number of Participants*Type of Model	0
Moving Average	0
Error Variance*Type of Model	0
Error Variance*Moving Average	0
Fixed Level	0
Number of Participants*Fixed Level	0
Number of Primary Studies*Type of Model	0
Fixed Level*Type of Model	0
Moving Average*Type of Model	0
Fixed Level*Moving Average	0
Number of Primary Studies*Fixed Level	0
Number of Primary Studies*Moving Average	0
Autocorrelation*Moving Average	0
Total Explained	0.9983

Table A18
Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Level-two Phase Effect

	η^2
Error Variance	0.4326
Number of Primary Studies	0.222
Number of Participants	0.12348
Type of Model	0.04868
Autocorrelation*Type of Model	0.03792
Number of Primary Studies*Error Variance	0.02838
Number of Participants*Error Variance	0.02444
Autocorrelation	0.01483
Series Length	0.00974
Number of Participants*Number of Primary Studies	0.00554
Series Length*Autocorrelation	0.00209
Series Length*Number of Participants	0.00202
Series Length*Error Variance	0.00128
Number of Primary Studies*Type of Model	0.00113
Series Length*Type of Model	0.00095
Number of Participants*Autocorrelation	0.00068
Number of Participants*Type of Model	0.00061
Error Variance*Type of Model	0.00048
Fixed Level	0.0004
Moving Average*Type of Model	0.00037
Fixed Level*Autocorrelation	0.00034
Error Variance*Autocorrelation	0.00033
Fixed Level*Type of Model	0.0002
Number of Primary Studies*Autocorrelation	0.00018
Error Variance*Moving Average	0.00015
Moving Average	0.00011
Number of Participants*Moving Average	0.00008
Fixed Level*Moving Average	0.00005
Series Length*Moving Average	0.00005
Number of Participants*Fixed Level	0.00004
Series Length*Fixed Level	0.00004
Number of Primary Studies*Moving Average	0.00004
Number of Primary Studies*Fixed Level	0.00002
Series Length*Number of Primary Studies	0.00001
Fixed Level*Error Variance	0
Autocorrelation*Moving Average	0
Total Explained	0.9593

Table A19

Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Level-two Interaction Effect

	η^2
Error Variance	0.34932
Series Length	0.25701
Number of Primary Studies	0.12644
Number of Participants	0.07658
Type of Model	0.03065
Number of Primary Studies*Error Variance	0.02454
Autocorrelation	0.02302
Autocorrelation*Type of Model	0.02286
Series Length*Error Variance	0.0201
Number of Participants*Error Variance	0.01497
Series Length*Type of Model	0.01144
Series Length*Number of Primary Studies	0.0082
Series Length*Number of Participants	0.00653
Series Length*Autocorrelation	0.00533
Number of Participants*Number of Primary Studies	0.00423
Error Variance*Autocorrelation	0.00133
Number of Primary Studies*Type of Model	0.00074
Error Variance*Type of Model	0.00041
Number of Participants*Type of Model	0.00033
Number of Participants*Autocorrelation	0.00011
Number of Primary Studies*Autocorrelation	0.00009
Fixed Level	0.00004
Number of Primary Studies*Fixed Level	0.00001
Number of Participants*Moving Average	0.00001
Moving Average	0.00001
Error Variance*Moving Average	0.00001
Fixed Level*Moving Average	0.00001
Fixed Level*Autocorrelation	0
Series Length*Moving Average	0
Moving Average*Type of Model	0
Series Length*Fixed Level	0
Fixed Level*Error Variance	0
Number of Participants*Fixed Level	0
Number of Primary Studies*Moving Average	0
Fixed Level*Type of Model	0
Autocorrelation*Moving Average	0
Total Explained	0.9843

Table A20

Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Level-one Variance

	η^2
Type of Model	0.20268
Error Variance	0.17296
Error Variance*Type of Model	0.10587
Autocorrelation*Type of Model	0.0703
Series Length	0.06046
Number of Primary Studies	0.05264
Error Variance*Autocorrelation*Type of Model	0.05095
Autocorrelation	0.04265
Error Variance*Autocorrelation	0.03552
Series Length*Autocorrelation	0.02984
Number of Participants	0.02013
Number of Primary Studies*Type of Model	0.01596
Number of Primary Studies*Error Variance	0.01467
Series Length*Error Variance	0.01364
Series Length*Error Variance*Autocorrelation	0.01362
Series Length*Autocorrelation*Type of Model	0.01164
Number of Primary Studies*Error Variance*Type of Model	0.00925
Series Length*Error Variance*Autocorrelation*Type of Model	0.00789
Number of Participants*Error Variance	0.00697
Number of Participants*Type of Model	0.00659
Series Length*Type of Model	0.00602
Series Length*Number of Primary Studies	0.00525
Series Length*Error Variance*Type of Model	0.00481
Number of Participants*Error Variance*Type of Model	0.004
Series Length*Number of Participants	0.00316
Number of Primary Studies*Autocorrelation	0.00266
Series Length*Number of Primary Studies*Autocorrelation	0.00259
Series Length*Number of Primary Studies*Type of Model	0.00203
Number of Primary Studies*Autocorrelation*Type of Model	0.00201
Series Length*Number of Primary Studies*Error Variance	0.00196
Series Length*Number of Primary Studies*Error Variance*Autocorrelation	0.00174
Number of Participants*Autocorrelation	0.00151
Number of Primary Studies*Error Variance*Autocorrelation	0.00146
Number of Participants*Number of Primary Studies	0.00104
Series Length*Number of Primary Studies*Autocorrelation*Type of Model	0.00103
Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.00103
Series Length*Number of Participants*Error Variance*Autocorrelation	0.00093
Number of Participants*Autocorrelation*Type of Model	0.00092
Series Length*Number of Participants*Autocorrelation	0.00086

Table A20 (Continued)

	η^2
Series Length*Number of Primary Studies*Error Variance*Type of Model	0.00083
Series Length*Number of Participants*Type of Model	0.00079
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation	0.00072
Series Length*Number of Participants*Error Variance	0.0007
Series Length*Number of Participants*Autocorrelation*Type of Model	0.00047
Number of Participants*Error Variance*Autocorrelation*Type of Model	0.00042
Number of Participants*Number of Primary Studies*Autocorrelation	0.00038
Number of Participants*Error Variance*Autocorrelation	0.00033
Number of Participants*Number of Primary Studies*Error Variance	0.00031
Number of Participants*Number of Primary Studies*Type of Model	0.0003
Series Length*Number of Participants*Error Variance*Type of Model	0.00027
Series Length*Number of Participants*Number of Primary Studies*Error Variance	0.00018
Series Length*Number of Participants*Number of Primary Studies*Type of Model	0.000155
Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.000147
Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.000145
Series Length*Number of Participants*Number of Primary Studies	0.000139
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	9.9E-05
Series Length*Number of Primary Studies*Fixed Level*Autocorrelation	9.09E-05
Series Length*Number of Participants*Fixed Level*Autocorrelation	8.57E-05
Number of Primary Studies*Fixed Level*Autocorrelation*Type of Model	5.22E-05
Number of Participants*Number of Primary Studies*Fixed Level*Error Variance	4.97E-05
Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	4.6E-05
Number of Participants*Moving Average	4.46E-05
Series Length*Fixed Level*Error Variance*Autocorrelation	4.16E-05
Number of Participants*Number of Primary Studies*Error Variance*Moving Average	4.15E-05
Series Length*Number of Primary Studies*Error Variance*Moving Average	4.03E-05
Number of Primary Studies*Fixed Level	3.87E-05
Series Length*Number of Participants*Moving Average	3.85E-05
Moving Average*Type of Model	3.83E-05
Series Length*Fixed Level*Error Variance*Moving Average	3.75E-05
Number of Participants*Number of Primary Studies*Fixed Level*Moving Average	3.75E-05
Number of Participants*Fixed Level*Autocorrelation*Type of Model	3.48E-05
Series Length*Error Variance*Moving Average	3.24E-05
Series Length*Number of Participants*Fixed Level	3.05E-05
Fixed Level*Error Variance*Autocorrelation*Type of Model	2.89E-05
Number of Participants*Number of Primary Studies*Moving Average	2.78E-05
Number of Primary Studies*Fixed Level*Autocorrelation	2.76E-05
Error Variance*Moving Average*Type of Model	2.67E-05
Series Length*Number of Primary Studies*Fixed Level	2.59E-05

Table A20(Continued)

	η^2
Series Length*Moving Average*Type of Model	2.42E-05
Series Length*Number of Primary Studies*Moving Average	2.38E-05
Moving Average	2.36E-05
Number of Participants*Fixed Level*Autocorrelation	2.34E-05
Number of Participants*Error Variance*Moving Average	2.28E-05
Fixed Level*Autocorrelation	2.17E-05
Number of Primary Studies*Error Variance*Moving Average	2.17E-05
Series Length*Number of Participants*Error Variance*Moving Average	2.14E-05
Number of Participants*Moving Average*Type of Model	1.99E-05
Fixed Level*Autocorrelation*Type of Model	1.9E-05
Series Length*Fixed Level*Moving Average	1.86E-05
Series Length*Error Variance*Moving Average*Type of Model	1.82E-05
Series Length*Number of Participants*Moving Average*Type of Model	1.73E-05
Number of Primary Studies*Moving Average	1.67E-05
Number of Primary Studies*Fixed Level*Type of Model	1.5E-05
Number of Participants*Number of Primary Studies*Moving Average*Type of Model	1.45E-05
Fixed Level*Error Variance*Moving Average	1.43E-05
Number of Primary Studies*Error Variance*Moving Average*Type of Model	1.39E-05
Series Length*Fixed Level*Autocorrelation	1.39E-05
Series Length*Number of Primary Studies*Fixed Level*Type of Model	1.36E-05
Series Length*Number of Primary Studies*Moving Average*Type of Model	1.35E-05
Series Length*Moving Average	1.25E-05
Series Length*Number of Participants*Fixed Level*Error Variance	1.22E-05
Number of Participants*Fixed Level*Error Variance*Type of Model	1.19E-05
Number of Participants*Fixed Level*Error Variance	1.17E-05
Fixed Level*Moving Average	1.16E-05
Fixed Level*Error Variance	1.14E-05
Fixed Level*Error Variance*Autocorrelation	1.09E-05
Number of Participants*Fixed Level*Error Variance*Moving Average	1.08E-05
Number of Primary Studies*Moving Average*Type of Model	1.08E-05
Series Length*Number of Participants*Fixed Level*Type of Model	1.06E-05
Number of Participants*Fixed Level*Error Variance*Autocorrelation	1.05E-05
Number of Participants*Error Variance*Moving Average*Type of Model	9.75E-06
Series Length*Fixed Level	9.74E-06
Series Length*Fixed Level*Moving Average*Type of Model	9E-06
Error Variance*Moving Average	8.83E-06
Series Length*Fixed Level*Autocorrelation*Type of Model	8.24E-06
Number of Participants*Number of Primary Studies*Fixed Level	7.59E-06
Number of Participants*Fixed Level*Type of Model	7.57E-06

Table A20 (Continued)

	η^2
Series Length*Fixed Level*Type of Model	7.47E-06
Number of Primary Studies*Fixed Level*Error Variance*Autocorrelation	7.4E-06
Fixed Level*Error Variance*Type of Model	7.23E-06
Fixed Level*Moving Average*Type of Model	7.19E-06
Fixed Level*Type of Model	6.63E-06
Number of Primary Studies*Fixed Level*Error Variance	6.33E-06
Number of Primary Studies*Fixed Level*Error Variance*Moving Average	5.74E-06
Series Length*Number of Primary Studies*Fixed Level*Error Variance	5.46E-06
Series Length*Fixed Level*Error Variance*Type of Model	5.01E-06
Number of Primary Studies*Fixed Level*Moving Average*Type of Model	3.96E-06
Fixed Level*Error Variance*Moving Average*Type of Model	3.82E-06
Number of Primary Studies*Fixed Level*Error Variance*Type of Model	3.6E-06
Number of Participants*Number of Primary Studies*Fixed Level*Type of Model	3.03E-06
Number of Participants*Fixed Level*Moving Average*Type of Model	2.48E-06
Fixed Level	1.7E-06
Number of Participants*Fixed Level*Moving Average	1.43E-06
Series Length*Number of Participants*Fixed Level*Moving Average	1.42E-06
Series Length*Fixed Level*Error Variance	1.16E-06
Number of Participants*Fixed Level	1.14E-06
Series Length*Number of Participants*Number of Primary Studies*Moving Average	7.57E-07
Series Length*Number of Participants*Number of Primary Studies*Fixed Level	4.96E-07
Number of Primary Studies*Fixed Level*Moving Average	2.88E-07
Series Length*Number of Primary Studies*Fixed Level*Moving Average	2.57E-07
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Series Length*Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Series Length*Number of Primary Studies*Autocorrelation*Moving Average	0
Number of Participants*Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Series Length*Error Variance*Autocorrelation*Moving Average	0
Number of Participants*Error Variance*Autocorrelation*Moving Average	0
Number of Primary Studies*Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Series Length*Fixed Level*Autocorrelation*Moving Average	0
Number of Participants*Fixed Level*Autocorrelation*Moving Average	0
Number of Primary Studies*Fixed Level*Autocorrelation*Moving Average	0
Fixed Level*Error Variance*Autocorrelation*Moving Average	0

Table A20 (Continued)

	η^2
Autocorrelation*Moving Average*Type of Model	0
Series Length*Autocorrelation*Moving Average*Type of Model	0
Number of Participants*Autocorrelation*Moving Average*Type of Model	0
Number of Primary Studies*Autocorrelation*Moving Average*Type of Model	0
Error Variance*Autocorrelation*Moving Average*Type of Model	0
Fixed Level*Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9972

Table A21

Eta-Squared Values (η^2) for the Association of the Design Factors with the RMSE Values for the Autocorrelation Parameter

	η^2
Type of Model	0.33264
Autocorrelation*Type of Model	0.24376
Autocorrelation	0.23321
Series Length	0.05116
Number of Primary Studies	0.04356
Series Length*Autocorrelation	0.01881
Number of Participants	0.01558
Number of Primary Studies*Autocorrelation	0.00809
Error Variance	0.0078
Number of Primary Studies*Type of Model	0.00714
Series Length*Type of Model	0.00521
Number of Participants*Autocorrelation	0.00416
Error Variance*Type of Model	0.00405
Number of Participants*Type of Model	0.00224
Series Length*Number of Primary Studies	0.00196
Series Length*Error Variance	0.00112
Series Length*Number of Participants	0.0007
Number of Participants*Number of Primary Studies	0.00047
Error Variance*Autocorrelation	0.00023
Number of Primary Studies*Error Variance	0.00018
Number of Participants*Error Variance	0.00007
Number of Primary Studies*Fixed Level	0.00001
Error Variance*Moving Average	0.00001
Fixed Level*Error Variance	0.00001
Series Length*Moving Average	0.00001
Fixed Level*Moving Average	0.00001
Moving Average	0.00001
Number of Participants*Fixed Level	0.00001

Table A21 (Continued)

	η^2
Number of Primary Studies*Moving Average	0
Fixed Level*Autocorrelation	0
Moving Average*Type of Model	0
Number of Participants*Moving Average	0
Series Length*Fixed Level	0
Fixed Level*Type of Model	0
Fixed Level	0
Autocorrelation*Moving Average	0
Total Explained	0.9822

Table A22

Eta-Squared Values (η^2) for the Association of the Design Factors with the Bias Estimates for the Moving Average Parameter

	η^2
Moving Average*Type of Model	0.45407
Autocorrelation	0.26246
Moving Average	0.0984
Autocorrelation*Type of Model	0.04353
Error Variance	0.02624
Type of Model	0.02417
Error Variance*Autocorrelation	0.02088
Error Variance*Type of Model	0.00983
Number of Primary Studies	0.00841
Series Length*Autocorrelation	0.00603
Series Length	0.00587
Number of Participants	0.00286
Number of Primary Studies*Type of Model	0.00264
Number of Primary Studies*Moving Average	0.00159
Series Length*Moving Average	0.00146
Number of Primary Studies*Error Variance	0.00123
Number of Participants*Type of Model	0.00099
Series Length*Type of Model	0.00063
Series Length*Error Variance	0.00059
Series Length*Number of Primary Studies	0.00057
Number of Participants*Error Variance	0.00054
Number of Primary Studies*Autocorrelation	0.00052
Number of Participants*Autocorrelation	0.00029
Number of Participants*Moving Average	0.00028
Error Variance*Moving Average	0.00024
Series Length*Number of Participants	0.00017

Table A22 (Continued)

	η^2
Number of Participants*Number of Primary Studies	0.00005
Fixed Level*Moving Average	0
Fixed Level*Autocorrelation	0
Number of Participants*Fixed Level	0
Fixed Level*Type of Model	0
Series Length*Fixed Level	0
Number of Primary Studies*Fixed Level	0
Fixed Level	0
Fixed Level*Error Variance	0
Autocorrelation*Moving Average	0
Total Explained	0.9745

Table A23

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Level-three for the Phase Effect

	η^2
Error Variance	0.3074
Number of Participants	0.2965
Number of Participants*Error Variance	0.09014
Number of Primary Studies	0.07271
Number of Primary Studies*Error Variance	0.03461
Series Length	0.03445
Number of Participants*Number of Primary Studies	0.03085
Series Length*Error Variance	0.02772
Autocorrelation	0.01033
Error Variance*Autocorrelation	0.00953
Type of Model	0.00715
Series Length*Number of Primary Studies	0.00713
Autocorrelation*Type of Model	0.00438
Error Variance*Type of Model	0.00372
Series Length*Number of Participants	0.00281
Number of Primary Studies*Moving Average	0.00246
Number of Primary Studies*Autocorrelation	0.00223
Number of Participants*Moving Average	0.00211
Number of Participants*Autocorrelation	0.00137
Fixed Level	0.00113
Series Length*Autocorrelation	0.00105
Moving Average	0.00086
Number of Participants*Type of Model	0.00072
Fixed Level*Error Variance	0.00063

Table A23 (Continued)

	η^2
Series Length*Moving Average	0.00036
Fixed Level*Moving Average	0.00032
Number of Primary Studies*Type of Model	0.00029
Series Length*Type of Model	0.00026
Fixed Level*Autocorrelation	0.00024
Error Variance*Moving Average	0.0001
Number of Primary Studies*Fixed Level	0.00006
Number of Participants*Fixed Level	0.00004
Moving Average*Type of Model	0.00003
Series Length*Fixed Level	0.00001
Fixed Level*Type of Model	0
Autocorrelation*Moving Average	0
Total Explained	0.9537

Table A24

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Level-three for the Interaction Effect

	η^2
Series Length	0.29368
Error Variance	0.14714
Number of Participants	0.14337
Series Length*Error Variance	0.11049
Series Length*Number of Participants	0.05523
Number of Primary Studies	0.03384
Number of Participants*Error Variance	0.03036
Series Length*Number of Primary Studies	0.02904
Autocorrelation	0.01867
Number of Participants*Number of Primary Studies	0.01787
Number of Primary Studies*Error Variance	0.01609
Type of Model	0.01449
Series Length*Autocorrelation	0.00932
Series Length*Type of Model	0.0075
Autocorrelation*Type of Model	0.00691
Error Variance*Type of Model	0.00606
Error Variance*Autocorrelation	0.00392
Number of Participants*Autocorrelation	0.00167
Number of Primary Studies*Autocorrelation	0.00148
Number of Primary Studies*Type of Model	0.00095
Fixed Level*Moving Average	0.00065
Fixed Level*Autocorrelation	0.00061

Table A24 (Continued)

	η^2
Number of Participants*Type of Model	0.00049
Series Length*Moving Average	0.00028
Number of Primary Studies*Fixed Level	0.00021
Series Length*Fixed Level	0.0002
Error Variance*Moving Average	0.00009
Number of Participants*Moving Average	0.00008
Number of Primary Studies*Moving Average	0.00006
Moving Average	0.00005
Number of Participants*Fixed Level	0.00005
Fixed Level	0.00004
Fixed Level*Error Variance	0.00002
Fixed Level*Type of Model	0.00001
Moving Average*Type of Model	0
Autocorrelation*Moving Average	0
Total Explained	0.9509

Table A25

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Level-two for the Phase Effect

	η^2
Type of Model	0.42088
Autocorrelation*Type of Model	0.22263
Error Variance*Type of Model	0.08454
Error Variance	0.05333
Autocorrelation	0.04453
Number of Primary Studies*Type of Model	0.03959
Error Variance*Autocorrelation*Type of Model	0.03639
Number of Participants*Type of Model	0.02367
Number of Primary Studies*Autocorrelation*Type of Model	0.01806
Number of Participants*Autocorrelation*Type of Model	0.01042
Error Variance*Autocorrelation	0.00864
Series Length*Type of Model	0.00707
Number of Primary Studies*Error Variance*Type of Model	0.00493
Series Length*Autocorrelation*Type of Model	0.00437
Number of Participants*Error Variance*Type of Model	0.0025
Number of Primary Studies*Autocorrelation	0.00236
Number of Primary Studies	0.0016
Number of Participants*Autocorrelation	0.00138
Series Length*Error Variance*Type of Model	0.00126
Number of Participants*Number of Primary Studies*Type of Model	0.00114

Table A25 (Continued)

	η^2
Number of Participants	0.00098
Series Length*Autocorrelation	0.00095
Number of Participants*Number of Primary Studies	0.00081
Series Length*Number of Primary Studies	0.00058
Series Length*Number of Participants	0.00029
Number of Primary Studies*Error Variance	0.00023
Number of Participants*Error Variance	0.00022
Series Length*Error Variance	0.0002
Series Length*Error Variance*Autocorrelation	0.00013
Series Length*Number of Primary Studies*Type of Model	0.00013
Number of Primary Studies*Error Variance*Autocorrelation	0.00011
Number of Participants*Number of Primary Studies*Autocorrelation	0.0001
Series Length*Number of Participants*Type of Model	0.00008
Number of Participants*Number of Primary Studies*Error Variance	0.00007
Number of Participants*Error Variance*Autocorrelation	0.00006
Series Length*Number of Participants*Error Variance	0.00004
Series Length*Number of Participants*Number of Primary Studies	0.00004
Series Length*Number of Participants*Autocorrelation	0.00002
Fixed Level*Moving Average	0.00002
Series Length*Number of Primary Studies*Autocorrelation	0.00002
Series Length*Number of Participants*Moving Average	0.00001
Series Length*Fixed Level*Autocorrelation	0.00001
Number of Primary Studies*Fixed Level*Moving Average	0.00001
Series Length	0.00001
Number of Primary Studies*Fixed Level*Autocorrelation	0.00001
Fixed Level*Error Variance*Autocorrelation	0.00001
Error Variance*Moving Average	0.00001
Number of Primary Studies*Error Variance*Moving Average	0.00001
Fixed Level*Error Variance*Moving Average	0.00001
Fixed Level	0.00001
Number of Participants*Fixed Level*Error Variance	0.00001
Series Length*Error Variance*Moving Average	6.28E-06
Number of Participants*Fixed Level	6.23E-06
Series Length*Moving Average	5.39E-06
Number of Primary Studies*Moving Average	5.25E-06
Fixed Level*Autocorrelation	5.23E-06
Series Length*Number of Primary Studies*Error Variance	4.32E-06
Moving Average*Type of Model	4.04E-06
Number of Participants*Number of Primary Studies*Fixed Level	3.99E-06
Number of Primary Studies*Moving Average*Type of Model	3.79E-06

Table A25(Continued)

	η^2
Moving Average	3.19E-06
Number of Primary Studies*Fixed Level	2.72E-06
Number of Participants*Number of Primary Studies*Moving Average	2.69E-06
Number of Participants*Fixed Level*Moving Average	2.18E-06
Number of Participants*Moving Average*Type of Model	2.06E-06
Error Variance*Moving Average*Type of Model	1.93E-06
Series Length*Fixed Level*Error Variance	1.69E-06
Series Length*Fixed Level*Moving Average	1.5E-06
Series Length*Number of Primary Studies*Fixed Level	1.22E-06
Number of Participants*Fixed Level*Autocorrelation	1.13E-06
Series Length*Moving Average*Type of Model	7E-07
Number of Participants*Error Variance*Moving Average	5.8E-07
Number of Participants*Fixed Level*Type of Model	4.6E-07
Number of Participants*Moving Average	4.3E-07
Fixed Level*Autocorrelation*Type of Model	3.99E-07
Fixed Level*Moving Average*Type of Model	3.8E-07
Series Length*Number of Primary Studies*Moving Average	3.03E-07
Number of Primary Studies*Fixed Level*Error Variance	2.58E-07
Series Length*Number of Participants*Fixed Level	2.38E-07
Series Length*Fixed Level*Type of Model	2.34E-07
Series Length*Fixed Level	2.18E-07
Number of Primary Studies*Fixed Level*Type of Model	1.95E-07
Fixed Level*Error Variance*Type of Model	1.94E-07
Fixed Level*Error Variance	5.3E-08
Fixed Level*Type of Model	5.1E-08
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9945

Table A26

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Level-two for the Interaction Effect

	η^2
Type of Model	0.38353
Autocorrelation*Type of Model	0.2222
Autocorrelation	0.0535
Series Length	0.05237
Number of Primary Studies*Type of Model	0.04095
Series Length*Type of Model	0.04015
Error Variance*Type of Model	0.03303
Error Variance	0.03015
Number of Participants*Type of Model	0.02338
Number of Primary Studies*Autocorrelation*Type of Model	0.02174
Error Variance*Autocorrelation*Type of Model	0.01654
Series Length*Autocorrelation*Type of Model	0.01482
Number of Participants*Autocorrelation*Type of Model	0.01208
Series Length*Autocorrelation	0.00728
Error Variance*Autocorrelation	0.00575
Series Length*Error Variance	0.00574
Series Length*Number of Primary Studies*Type of Model	0.00358
Number of Primary Studies*Error Variance*Type of Model	0.00251
Series Length*Error Variance*Type of Model	0.00241
Number of Primary Studies*Autocorrelation	0.0022
Series Length*Number of Participants*Type of Model	0.00201
Number of Participants*Number of Primary Studies*Type of Model	0.00141
Series Length*Number of Primary Studies	0.0014
Number of Participants*Autocorrelation	0.00135
Number of Participants*Error Variance*Type of Model	0.00134
Series Length*Number of Participants	0.00109
Number of Primary Studies	0.00089
Number of Participants*Number of Primary Studies	0.00086
Series Length*Number of Primary Studies*Error Variance	0.00076
Series Length*Number of Participants*Error Variance	0.00068
Number of Participants	0.00046
Series Length*Error Variance*Autocorrelation	0.00038
Number of Primary Studies*Error Variance	0.00023
Number of Participants*Error Variance	0.00019
Number of Participants*Number of Primary Studies*Error Variance	0.00011
Series Length*Number of Participants*Number of Primary Studies	0.00009
Series Length*Number of Primary Studies*Autocorrelation	0.00006
Number of Participants*Number of Primary Studies*Autocorrelation	0.00006
Fixed Level*Error Variance	0.00005

Table A26 (Continued)

	η^2
Series Length*Number of Participants*Autocorrelation	0.00005
Number of Primary Studies*Error Variance*Autocorrelation	0.00003
Error Variance*Moving Average	0.00003
Number of Participants*Number of Primary Studies*Moving Average	0.00003
Number of Participants*Error Variance*Autocorrelation	0.00002
Number of Participants*Number of Primary Studies*Fixed Level	0.00002
Series Length*Fixed Level*Autocorrelation	0.00001
Series Length*Error Variance*Moving Average	0.00001
Number of Primary Studies*Fixed Level*Autocorrelation	0.00001
Number of Participants*Fixed Level*Moving Average	0.00001
Number of Primary Studies*Error Variance*Moving Average	0.00001
Series Length*Number of Participants*Moving Average	0.00001
Number of Participants*Error Variance*Moving Average	8.08E-06
Fixed Level*Moving Average*Type of Model	7.82E-06
Fixed Level*Error Variance*Moving Average	7.72E-06
Number of Participants*Fixed Level*Error Variance	6.75E-06
Fixed Level*Moving Average	6.38E-06
Fixed Level	6.28E-06
Fixed Level*Error Variance*Autocorrelation	5.68E-06
Fixed Level*Autocorrelation	5.61E-06
Number of Participants*Fixed Level*Autocorrelation	5.37E-06
Series Length*Number of Primary Studies*Moving Average	5.15E-06
Number of Primary Studies*Fixed Level*Moving Average	4.6E-06
Number of Participants*Moving Average*Type of Model	4.38E-06
Moving Average	4.31E-06
Series Length*Fixed Level*Error Variance	4.19E-06
Series Length*Number of Primary Studies*Fixed Level	3.4E-06
Series Length*Fixed Level*Type of Model	3.01E-06
Fixed Level*Autocorrelation*Type of Model	2.98E-06
Series Length*Moving Average	2.94E-06
Number of Participants*Moving Average	2.83E-06
Series Length*Moving Average*Type of Model	2.73E-06
Number of Primary Studies*Moving Average	2.31E-06
Number of Primary Studies*Fixed Level*Type of Model	2.11E-06
Error Variance*Moving Average*Type of Model	1.2E-06
Number of Primary Studies*Fixed Level	1.15E-06
Moving Average*Type of Model	8.32E-07
Series Length*Number of Participants*Fixed Level	4.9E-07
Fixed Level*Type of Model	4.78E-07
Number of Primary Studies*Moving Average*Type of Model	4.73E-07

Table A26 (Continued)

	η^2
Number of Primary Studies*Fixed Level*Error Variance	4.29E-07
Fixed Level*Error Variance*Type of Model	3.97E-07
Series Length*Fixed Level*Moving Average	3.51E-07
Number of Participants*Fixed Level*Type of Model	2.18E-07
Number of Participants*Fixed Level	2E-08
Series Length*Fixed Level	1.3E-08
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9877

Table A27

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Level-one Variance

	η^2
Autocorrelation	0.53353
Series Length*Type of Model	0.07255
Number of Primary Studies	0.06228
Series Length*Autocorrelation	0.05339
Series Length	0.04854
Series Length*Autocorrelation*Type of Model	0.04281
Number of Primary Studies*Autocorrelation	0.04087
Autocorrelation*Type of Model	0.02892
Number of Participants	0.02502
Type of Model	0.0211
Number of Participants*Autocorrelation	0.01566
Number of Primary Studies*Autocorrelation*Type of Model	0.00709
Number of Primary Studies*Type of Model	0.00423
Error Variance*Autocorrelation	0.00419
Error Variance*Autocorrelation*Type of Model	0.00275
Series Length*Number of Primary Studies	0.00272
Number of Participants*Autocorrelation*Type of Model	0.00239
Error Variance*Type of Model	0.00236
Series Length*Number of Participants*Number of Primary Studies	0.00192
Number of Participants*Number of Primary Studies*Autocorrelation	0.0018
Series Length*Error Variance*Autocorrelation	0.00141

Table A27(Continued)

	η^2
Series Length*Number of Primary Studies*Type of Model	0.0014
Number of Participants*Type of Model	0.00127
Series Length*Number of Primary Studies*Autocorrelation	0.00114
Series Length*Number of Participants	0.00075
Series Length*Number of Primary Studies*Error Variance	0.00072
Series Length*Number of Participants*Autocorrelation	0.00071
Series Length*Error Variance	0.00063
Error Variance	0.00043
Series Length*Number of Participants*Type of Model	0.00036
Number of Participants*Number of Primary Studies	0.00035
Number of Primary Studies*Error Variance	0.00033
Number of Primary Studies*Error Variance*Type of Model	0.00022
Series Length*Number of Participants*Error Variance	0.00019
Series Length*Error Variance*Type of Model	0.00018
Number of Participants*Error Variance*Type of Model	0.00014
Number of Participants*Error Variance	0.00013
Number of Participants*Number of Primary Studies*Type of Model	0.00003
Number of Participants*Fixed Level	0.00002
Number of Primary Studies*Error Variance*Autocorrelation	0.00001
Number of Participants*Error Variance*Autocorrelation	0.00001
Fixed Level*Error Variance*Moving Average	0.00001
Number of Participants*Number of Primary Studies*Error Variance	0.00001
Number of Participants*Fixed Level*Type of Model	0
Series Length*Number of Participants*Fixed Level	0
Series Length*Fixed Level*Autocorrelation	0
Series Length*Error Variance*Moving Average	0
Number of Participants*Error Variance*Moving Average	0
Fixed Level*Autocorrelation*Type of Model	0
Fixed Level*Autocorrelation	0
Error Variance*Moving Average*Type of Model	0
Number of Primary Studies*Error Variance*Moving Average	2.23E-06
Number of Participants*Fixed Level*Autocorrelation	2.12E-06
Number of Participants*Moving Average	1.8E-06
Series Length*Moving Average*Type of Model	1.63E-06
Series Length*Fixed Level*Moving Average	1.45E-06
Number of Primary Studies*Moving Average	1.4E-06
Series Length*Number of Primary Studies*Fixed Level	1.38E-06
Number of Primary Studies*Fixed Level*Moving Average	1.29E-06
Series Length*Moving Average	1.12E-06
Error Variance*Moving Average	9.63E-07

Table A27(Continued)

	η^2
Fixed Level*Moving Average	7.96E-07
Series Length*Fixed Level*Error Variance	7.65E-07
Moving Average*Type of Model	7.45E-07
Number of Primary Studies*Moving Average*Type of Model	7.22E-07
Number of Participants*Moving Average*Type of Model	6.29E-07
Fixed Level	6.12E-07
Fixed Level*Moving Average*Type of Model	5.7E-07
Number of Primary Studies*Fixed Level*Error Variance	4.68E-07
Number of Primary Studies*Fixed Level*Type of Model	4.04E-07
Series Length*Fixed Level	4E-07
Series Length*Number of Primary Studies*Moving Average	3.87E-07
Fixed Level*Type of Model	3.86E-07
Fixed Level*Error Variance*Autocorrelation	2.82E-07
Number of Participants*Number of Primary Studies*Fixed Level	2.25E-07
Number of Participants*Number of Primary Studies*Moving Average	1.93E-07
Moving Average	1.9E-07
Series Length*Number of Participants*Moving Average	1.62E-07
Series Length*Fixed Level*Type of Model	1.61E-07
Number of Participants*Fixed Level*Moving Average	1.16E-07
Number of Participants*Fixed Level*Error Variance	1.08E-07
Fixed Level*Error Variance	1.01E-07
Fixed Level*Error Variance*Type of Model	5.5E-08
Number of Primary Studies*Fixed Level	5.2E-08
Number of Primary Studies*Fixed Level*Autocorrelation	1E-09
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9846

Table A28

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI coverage for the Autocorrelation Parameter

	η^2
Autocorrelation*Type of Model	0.32371
Type of Model	0.16802
Autocorrelation	0.05901
Series Length*Autocorrelation	0.0576
Number of Primary Studies	0.05538
Series Length	0.04776
Series Length*Autocorrelation*Type of Model	0.03759
Number of Primary Studies*Autocorrelation	0.02618
Number of Participants	0.02514
Series Length*Number of Primary Studies	0.02123
Series Length*Type of Model	0.01438
Number of Participants*Autocorrelation	0.01396
Error Variance*Autocorrelation*Type of Model	0.01157
Series Length*Number of Participants	0.01117
Series Length*Number of Primary Studies*Autocorrelation	0.0111
Error Variance*Type of Model	0.01099
Number of Primary Studies*Type of Model	0.01022
Number of Participants*Number of Primary Studies	0.00758
Number of Primary Studies*Autocorrelation*Type of Model	0.00736
Error Variance	0.00692
Number of Participants*Number of Primary Studies*Autocorrelation	0.00587
Series Length*Number of Primary Studies*Type of Model	0.00497
Series Length*Number of Participants*Autocorrelation	0.00456
Number of Participants*Type of Model	0.00415
Series Length*Number of Participants*Number of Primary Studies	0.0041
Number of Participants*Autocorrelation*Type of Model	0.00362
Series Length*Number of Participants*Type of Model	0.00254
Series Length*Error Variance	0.0021
Number of Participants*Number of Primary Studies*Type of Model	0.002
Number of Primary Studies*Error Variance	0.00127
Error Variance*Autocorrelation	0.00124
Series Length*Error Variance*Autocorrelation	0.00108
Series Length*Number of Primary Studies*Error Variance	0.0007
Number of Primary Studies*Error Variance*Type of Model	0.00065
Number of Primary Studies*Error Variance*Autocorrelation	0.00063
Series Length*Error Variance*Type of Model	0.00051
Number of Participants*Error Variance	0.00037
Number of Participants*Error Variance*Type of Model	0.00034
Series Length*Number of Participants*Error Variance	0.00031

Table A28 (Continued)

	η^2
Number of Primary Studies*Moving Average	0.0001
Number of Primary Studies*Error Variance*Moving Average	0.0001
Number of Participants*Error Variance*Autocorrelation	0.00008
Number of Participants*Number of Primary Studies*Error Variance	0.00007
Fixed Level*Autocorrelation	0.00004
Fixed Level*Moving Average	0.00004
Error Variance*Moving Average	0.00003
Number of Participants*Moving Average	0.00003
Series Length*Error Variance*Moving Average	0.00003
Number of Participants*Number of Primary Studies*Fixed Level	0.00003
Number of Primary Studies*Fixed Level*Autocorrelation	0.00002
Number of Participants*Error Variance*Moving Average	0.00002
Series Length*Fixed Level*Moving Average	2.23E-05
Series Length*Number of Participants*Fixed Level	2.07E-05
Number of Participants*Fixed Level*Autocorrelation	1.75E-05
Series Length*Fixed Level*Autocorrelation	1.71E-05
Fixed Level*Error Variance*Type of Model	1.66E-05
Number of Participants*Fixed Level	1.49E-05
Series Length*Number of Primary Studies*Moving Average	1.45E-05
Number of Participants*Fixed Level*Moving Average	1.36E-05
Moving Average	1.26E-05
Fixed Level	1.21E-05
Error Variance*Moving Average*Type of Model	1.13E-05
Series Length*Fixed Level	1.12E-05
Number of Primary Studies*Fixed Level*Moving Average	1.08E-05
Moving Average*Type of Model	9.48E-06
Series Length*Moving Average*Type of Model	9.27E-06
Fixed Level*Type of Model	8.95E-06
Fixed Level*Autocorrelation*Type of Model	7.64E-06
Number of Participants*Number of Primary Studies*Moving Average	7.61E-06
Number of Participants*Moving Average*Type of Model	7.37E-06
Series Length*Fixed Level*Error Variance	7.04E-06
Number of Primary Studies*Moving Average*Type of Model	5.33E-06
Series Length*Moving Average	5.29E-06
Number of Participants*Fixed Level*Type of Model	3.44E-06
Fixed Level*Moving Average*Type of Model	3.04E-06
Number of Primary Studies*Fixed Level*Error Variance	2.19E-06
Fixed Level*Error Variance*Autocorrelation	1.95E-06
Number of Primary Studies*Fixed Level*Type of Model	1.86E-06
Series Length*Fixed Level*Type of Model	1.85E-06

Table A28 (Continued)

	η^2
Series Length*Number of Participants*Moving Average	1.83E-06
Series Length*Number of Primary Studies*Fixed Level	1.16E-06
Fixed Level*Error Variance*Moving Average	5.95E-07
Number of Participants*Fixed Level*Error Variance	2.9E-07
Fixed Level*Error Variance	1.26E-07
Number of Primary Studies*Fixed Level	7.1E-08
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9688

Table A29

Eta-Squared Values (η^2)for the Association of the Design Factors with the CI coverage for the Moving Average Parameter

	η^2
Moving Average*Type of Model	0.53854
Moving Average	0.2604
Autocorrelation	0.18233
Autocorrelation*Type of Model	0.005
Type of Model	0.00144
Series Length*Type of Model	0.00067
Series Length*Moving Average	0.00045
Error Variance*Autocorrelation	0.00043
Series Length*Autocorrelation	0.00033
Error Variance*Moving Average	0.0003
Error Variance*Type of Model	0.00016
Number of Primary Studies*Moving Average	0.00014
Error Variance	0.00013
Number of Primary Studies*Autocorrelation	0.00011
Number of Participants*Moving Average	0.00008
Number of Primary Studies*Type of Model	0.00008
Number of Primary Studies	0.00005
Number of Participants*Type of Model	0.00004
Number of Participants	0.00003
Number of Participants*Autocorrelation	0.00001
Number of Primary Studies*Error Variance	0.00001

Table A29 (Continued)

	η^2
Series Length*Error Variance	0.00001
Number of Participants*Error Variance	0
Fixed Level*Moving Average	0
Fixed Level*Autocorrelation	0
Series Length	0
Series Length*Number of Participants	0
Number of Participants*Number of Primary Studies	0
Series Length*Number of Primary Studies	0
Fixed Level	0
Fixed Level*Type of Model	0
Number of Participants*Fixed Level	0
Fixed Level*Error Variance	0
Series Length*Fixed Level	0
Number of Primary Studies*Fixed Level	0
Autocorrelation*Moving Average	0
Total Explained	0.9907

Table A30

Eta-Squared Values (η^2)for the Association of the Design Factors with the CI Width for the Level-one Variance

	η^2
Type of Model	0.17801
Series Length	0.15039
Number of Primary Studies	0.13942
Autocorrelation*Type of Model	0.08227
Number of Participants	0.06393
Autocorrelation	0.06279
Series Length*Type of Model	0.05071
Series Length*Autocorrelation*Type of Model	0.02111
Number of Primary Studies*Type of Model	0.01973
Series Length*Number of Primary Studies	0.01684
Series Length*Autocorrelation	0.01413
Number of Participants*Type of Model	0.01071
Number of Primary Studies*Autocorrelation*Type of Model	0.01065
Error Variance	0.00926
Series Length*Number of Participants	0.0091
Error Variance*Type of Model	0.00888
Number of Primary Studies*Autocorrelation	0.0088
Number of Participants*Number of Primary Studies	0.00785
Series Length*Number of Primary Studies*Type of Model	0.00724

Table A30(Continued)

	η^2
Series Length*Error Variance	0.00677
Series Length*Error Variance*Type of Model	0.00637
Number of Participants*Autocorrelation*Type of Model	0.00616
Error Variance*Autocorrelation*Type of Model	0.00533
Number of Participants*Autocorrelation	0.0052
Error Variance*Autocorrelation	0.00447
Series Length*Number of Participants*Type of Model	0.00444
Series Length*Error Variance*Autocorrelation	0.00341
Series Length*Number of Primary Studies*Autocorrelation	0.0032
Number of Primary Studies*Error Variance	0.00263
Number of Primary Studies*Error Variance*Type of Model	0.00254
Number of Participants*Number of Primary Studies*Type of Model	0.00227
Series Length*Number of Participants*Autocorrelation	0.00225
Series Length*Number of Primary Studies*Error Variance	0.00215
Series Length*Number of Participants*Number of Primary Studies	0.002
Number of Primary Studies*Error Variance*Autocorrelation	0.00186
Number of Participants*Error Variance	0.00176
Number of Participants*Error Variance*Type of Model	0.0017
Number of Participants*Number of Primary Studies*Autocorrelation	0.00162
Series Length*Number of Participants*Error Variance	0.00152
Number of Participants*Error Variance*Autocorrelation	0.0014
Number of Participants*Number of Primary Studies*Error Variance	0.00098
Number of Participants*Error Variance*Moving Average	0.00043
Series Length*Number of Primary Studies*Moving Average	0.00043
Number of Primary Studies*Error Variance*Moving Average	0.00042
Series Length*Moving Average	0.00042
Series Length*Moving Average*Type of Model	0.00042
Number of Participants*Number of Primary Studies*Moving Average	0.00042
Series Length*Number of Participants*Moving Average	0.00041
Number of Primary Studies*Moving Average	0.00041
Moving Average	0.00041
Moving Average*Type of Model	0.00041
Number of Primary Studies*Moving Average*Type of Model	0.000412
Number of Participants*Moving Average	0.000409
Number of Participants*Moving Average*Type of Model	0.000408
Series Length*Error Variance*Moving Average	0.000408
Error Variance*Moving Average*Type of Model	0.000403
Error Variance*Moving Average	0.000402
Number of Participants*Fixed Level*Moving Average	0.000149
Fixed Level*Error Variance*Moving Average	0.000142

Table A30 (Continued)

	η^2
Fixed Level*Moving Average	0.000127
Number of Primary Studies*Fixed Level*Moving Average	0.000126
Fixed Level*Moving Average*Type of Model	0.000125
Series Length*Fixed Level*Moving Average	0.000123
Fixed Level*Autocorrelation	9.91E-05
Number of Primary Studies*Fixed Level*Autocorrelation	9.87E-05
Fixed Level*Autocorrelation*Type of Model	9.78E-05
Series Length*Fixed Level*Autocorrelation	9.56E-05
Number of Participants*Fixed Level*Autocorrelation	8.91E-05
Fixed Level*Error Variance*Autocorrelation	8.55E-05
Number of Participants*Number of Primary Studies*Fixed Level	6.98E-05
Series Length*Fixed Level*Error Variance	6.91E-05
Series Length*Fixed Level	6.85E-05
Series Length*Number of Primary Studies*Fixed Level	6.85E-05
Number of Primary Studies*Fixed Level	6.75E-05
Series Length*Fixed Level*Type of Model	6.72E-05
Fixed Level	6.7E-05
Fixed Level*Error Variance	6.67E-05
Number of Primary Studies*Fixed Level*Type of Model	6.59E-05
Fixed Level*Type of Model	6.59E-05
Series Length*Number of Participants*Fixed Level	6.54E-05
Number of Primary Studies*Fixed Level*Error Variance	6.51E-05
Fixed Level*Error Variance*Type of Model	6.5E-05
Number of Participants*Fixed Level*Type of Model	6.45E-05
Number of Participants*Fixed Level	6.38E-05
Number of Participants*Fixed Level*Error Variance	5.83E-05
Autocorrelation*Moving Average	0
Series Length*Autocorrelation*Moving Average	0
Number of Participants*Autocorrelation*Moving Average	0
Number of Primary Studies*Autocorrelation*Moving Average	0
Error Variance*Autocorrelation*Moving Average	0
Fixed Level*Autocorrelation*Moving Average	0
Autocorrelation*Moving Average*Type of Model	0
Total Explained	0.9509

Table A31

Eta-Squared Values (η^2) for the Association of the Design Factors with the CI Width for the Autocorrelation Parameter

	η^2
Type of Model	0.54456
Autocorrelation*Type of Model	0.30678
Series Length	0.04354
Series Length*Type of Model	0.02829
Number of Primary Studies	0.02119
Autocorrelation	0.00992
Number of Participants	0.00834
Number of Primary Studies*Type of Model	0.00476
Number of Primary Studies*Autocorrelation	0.00328
Error Variance*Autocorrelation	0.00315
Number of Participants*Type of Model	0.00213
Number of Participants*Autocorrelation	0.00143
Error Variance*Type of Model	0.00065
Series Length*Number of Primary Studies	0.00057
Series Length*Autocorrelation	0.00044
Series Length*Error Variance	0.0004
Series Length*Number of Participants	0.00028
Number of Participants*Number of Primary Studies	0.00013
Error Variance	0.00009
Number of Primary Studies*Error Variance	0.00002
Number of Participants*Error Variance	0.00001
Fixed Level*Error Variance	0
Fixed Level*Type of Model	0
Fixed Level	0
Number of Participants*Moving Average	0
Series Length*Fixed Level	0
Number of Primary Studies*Moving Average	0
Fixed Level*Autocorrelation	0
Fixed Level*Moving Average	0
Moving Average*Type of Model	0
Series Length*Moving Average	0
Moving Average	0
Number of Participants*Fixed Level	0
Error Variance*Moving Average	0
Number of Primary Studies*Fixed Level	0
Autocorrelation*Moving Average	0
Total Explained	0.9800

Table A32

Eta-Squared Values (η²)for the Association of the Design Factors with the CI Width for the Moving Average Parameter

	η^2
Autocorrelation*Type of Model	0.27967
Type of Model	0.12097
Autocorrelation	0.06956
Error Variance*Autocorrelation*Type of Model	0.05626
Error Variance*Type of Model	0.02664
Number of Participants*fix*Error Variance*Autocorrelation*Type of Model	0.02572
Number of Primary Studies*Autocorrelation*Type of Model	0.02324
Series Length*Autocorrelation*Type of Model	0.02239
Series Length*Number of Participants*Error Variance*Autocorrelation*Type of Model	0.01625
Error Variance*Autocorrelation	0.01405
Number of Participants*Fixed Level*Error Variance*Type of Model	0.00919
Fixed Level*Error Variance*Autocorrelation*Type of Model	0.00916
Series Length*Number of Primary Studies*Error Variance*Autocorrelation*Model	0.00904

Table 33A Continued

Eta-Squared Values (η²)for the Association of the Design Factors with the CI Width for the Moving Average Parameter

Number of Primary Studies*Error Variance*Moving Average*Type of Model	0.00826
Series Length*Number of Primary Studies*Error Variance*the*Type of Model	0.00816
Series Length*fix*Error Variance*Autocorrelation*Type of Model	0.00749
Number of Participants*Moving Average*Type of Model	0.00743
Series Length*Number of Participants*Moving Average*Type of Model	0.00724
Series Length*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.00722
Series Length*Number of Participants*Number of Primary Studies*the*Type of Model	0.00703
Series Length*Error Variance*Moving Average*Type of Model	0.007
Number of Primary Studies*fix*Error Variance*Autocorrelation*Type of Model	0.00696
Error Variance*Moving Average*Type of Model	0.0069
Number of Participants*Number of Primary Studies*Moving Average*Type of Model	0.00684
Number of Participants*Number of Primary Studies*fix*Autocorrelation*Type of Model	0.00684
Error Variance	0.00671
Number of Participants*Fixed Level*Error Variance*Autocorrelation	0.00643
Series Length*Number of Primary Studies*Type of Model	0.00594
Number of Primary Studies*Autocorrelation	0.00584
Series Length*Autocorrelation	0.00566
Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.00565
Number of Participants*Error Variance*Autocorrelation*Type of Model	0.00497
Series Length*Number of Participants*Type of Model	0.00496
Series Length*Number of Participants*Autocorrelation*Type of Model	0.00474

Table A32 (Continued)

	η^2
Series Length*Number of Participants*fix*Autocorrelation*Type of Model	0.00457
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation*Type of Model	0.00443
Series Length*Number of Primary Studies*Autocorrelation*Type of Model	0.00412
Series Length*Number of Participants*Error Variance*Autocorrelation	0.00406
Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.00401
Number of Participants*Type of Model	0.00397
Series Length*Error Variance*Autocorrelation*Type of Model	0.00298
Number of Participants*Autocorrelation*Type of Model	0.00293
Number of Participants*Error Variance*Type of Model	0.00292
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.00285
Series Length*Number of Primary Studies*Moving Average*Type of Model	0.00274
Number of Primary Studies*Moving Average*Type of Model	0.00262
Series Length*Number of Participants*Error Variance*the*Type of Model	0.00248
Number of Participants*Error Variance*Moving Average*Type of Model	0.00242
Number of Participants*Fixed Level*Error Variance	0.0023
Fixed Level*Error Variance*Autocorrelation	0.00229
Series Length*Number of Primary Studies*Error Variance*Autocorrelation	0.00225
Number of Participants*Number of Primary Studies*Error Variance*the*Type of Model	0.002249
Series Length*Number of Participants*Fixed Level*Type of Model	0.002246
Series Length*Fixed Level*Autocorrelation*Type of Model	0.002228
Number of Primary Studies*Error Variance*Moving Average	0.002067
Series Length*Number of Primary Studies*Error Variance*Moving Average	0.00204
Moving Average*Type of Model	0.002029
Series Length*Moving Average*Type of Model	0.001928
Series Length*Fixed Level*Error Variance*Autocorrelation	0.001874
Number of Participants*Moving Average	0.001857
Series Length*Number of Participants*Moving Average	0.001809
Series Length*Number of Primary Studies*Fixed Level*Autocorrelation	0.001806
Series Length*Number of Participants*Number of Primary Studies*Moving Average	0.001758
Series Length*Error Variance*Moving Average	0.001748
Number of Primary Studies*Fixed Level*Error Variance*Autocorrelation	0.001739
Error Variance*Moving Average	0.001725
Number of Participants*Number of Primary Studies*Moving Average	0.00171
Number of Participants*Number of Primary Studies*Fixed Level*Autocorrelation	0.001709
Series Length*Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.001703
Number of Participants*Number of Primary Studies*fix*Error Variance*Type of Model	0.001695
Number of Participants*Number of Primary Studies*Autocorrelation*Type of Model	0.001658

Table A32 (Continued)

Series Length*Number of Primary Studies*fix*Error Variance*Type of Model	0.001642
Fixed Level*Error Variance*Type of Model	0.0015
Series Length*Number of Primary Studies	0.001489
Number of Primary Studies*Error Variance*Autocorrelation	0.001411
Series Length*Number of Participants*Number of Primary Studies*Type of Model	0.001393
Number of Participants*Number of Primary Studies*Type of Model	0.001274
Series Length*Fixed Level*Error Variance*Type of Model	0.001256
Number of Participants*Error Variance*Autocorrelation	0.001241
Series Length*Number of Participants	0.00124
Series Length*Number of Participants*Number of Primary Studies*fix*Error Variance	0.001231
Number of Primary Studies*Fixed Level*Autocorrelation*Type of Model	0.001212
Series Length*Number of Participants*Autocorrelation	0.001182
Series Length*Number of Participants*Fixed Level*Autocorrelation	0.001143
Number of Primary Studies*Fixed Level*Type of Model	0.001135
Number of Participants*Number of Primary Studies*Error Variance*Autocorrelation	0.001108
Number of Participants*Fixed Level*Autocorrelation*Type of Model	0.00109
Number of Participants*Number of Primary Studies*Fixed Level*Type of Model	0.001071
Number of Participants	0.001025
Series Length*Number of Primary Studies*Autocorrelation	0.001023
Number of Primary Studies*Fixed Level*Error Variance*Type of Model	0.001018
Number of Participants*Number of Primary Studies*Error Variance	0.001
Series Length*Number of Primary Studies*fix*Error Variance*Autocorrelation	0.000906
Series Length*Number of Participants*Number of Primary Studies*Error Variance*Type of Model	0.000844
Series Length*Error Variance*Type of Model	0.000765
Series Length*Error Variance*Autocorrelation	0.000749
Number of Participants*Autocorrelation	0.000729
Number of Participants*Error Variance	0.000728
Series Length*Number of Primary Studies*Moving Average	0.000685
Number of Primary Studies*Moving Average	0.000656
Series Length*Number of Participants*Error Variance*Moving Average	0.00062
Number of Participants*Error Variance*Moving Average	0.000606
Total Explained	0.9449