EFFECT SIZES FOR SINGLE CASE EXPERIMENTAL DESIGNS AND THEIR UTILITY FOR A META-ANALYSIS: A SIMULATION STUDY

JOO ANN LEE

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

SEPTEMBER 2015

ABSTRACT

There has been a lack of consensus as to the optimal effect size for use in meta-analyses involving Single Case Experimental Designs (SCEDs). SCEDs are a set of experimental designs which produce data akin to short interrupted time-series, where observations may not be independent due to autocorrelation. This thesis evaluated the statistical properties of various effect sizes for a reversal $\mathrm{ABA'B'}$ SCED via a simulation study. Hedges, Pustejovsky, and Shadish's (2012) Standardized Mean Difference effect size (δ_{HPS}) performed best when small to moderate degrees of autocorrelation were present. Partial regression coefficients also performed relatively well in most situations. The results recommend the utilization of δ_{HPS}: besides its favorable performance, δ_{HPS} has also been designed to be comparable to group-based effect sizes (Cohen's d) thus enabling the amalgamation of both SCEDs and group designs in a meta-analysis. Partial regression coefficients may also be used effectively in a meta-analysis of results from SCEDs.

ACKNOWLEDGEMENT

Thank you, first and foremost to my advisor, Dr. David Flora, for his patient guidance in supporting me to the completion of my M.A. thesis and advancing my understanding in various quantitative methods research. I would also like to express my deep gratitude to my thesis committee member, Dr. Jolynn Pek, for all her insights and hard questions which helped me further my own research, in addition to completing my M.A. thesis. Thank you also to Dr. Robert Cribbie and Dr. Michael Rotondi for all your ideas and feedback.

Thank you, as well to all my friends especially Cathy Labrish, Phil Chalmers, Hugh McCague, Victoria Ng, Alyssa Counsell, and Ryan Barnhart for all the support and suggestions along the way. The learning experience would not have been the same without you. Last but not least, I thank my loving family and husband for all their sacrifices made in supporting my aspiration to complete the M.A. degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Effect sizes for SCEDs and their utility for a meta-analysis: A simulation study

There is a need for review and summary of empirical research when there is a multitude of research effort and published findings on a given topic. Research synthesis is a structured review of research studies that incorporates various measures to guard against potential biases in the literature review (Chalmers, Hedges, & Cooper, 2002). Meta-analysis is the quantitative component in research synthesis and is used to minimize statistical imprecision by combining quantitative results from different studies (Chalmers et al., 2002). The major motivating context for this thesis is research synthesis and meta-analysis of data from single-case experimental designs (SCEDs; also known as *n*-of-1 trials, small-*n* designs, single-subject designs, intra-subject designs, single-case designs, single-system design, and interrupted time-series experimental designs, among others). More specifically, with the goal of finding suitable SCED effect size metrics for meta-analyses, this thesis evaluated various effect size statistics for a common type of SCED in terms of their performance across a variety of simulated short interrupted time-series data that exhibited autocorrelation and different types of effect. The possibility of an amalgamation of both SCEDs and group designs in a meta-analysis is also discussed, with applications primarily in the field of psychology.

SCEDs are a set of quasi-experimental designs that employ repeated data collection over time on a single unit of interest (e.g. Barlow, Nock, & Hersen, 2008, Gast & Ledford, 2014; Kazdin, 2011). The unit of interest may be an individual or a group such as a family or an organization, and SCEDs are frequently used to evaluate the effect of a treatment or intervention on the unit with the unit serving as its own control. Therefore, SCEDs are especially beneficial when the research areas studied have high variability or a low prevalence rate. The amount of research using SCEDs is growing and has produced important empirical findings in fields such as clinical psychology, occupation therapy, counseling, special education, social work, group dynamics, sports psychology, and medicine (e.g. Allen, Friman, & Sanger, 1992; Barker et al., 2013; Horner et al., 2005; Johnston & Smith, 2010; Macgowan & Wong, 2014; Perdices & Tate, 2009).

SCEDs refer to a set of experimental designs, rather than a single design, which generally have the following five basic characteristics in common (Allen et al., 1992):

1) There are repeated measurements of the outcome variable over time for each unit.
2) There must be a well-established baseline for each unit before entering the treatment or intervention phase. The baseline phase is important because it serves as the initial data against which subsequent data will be compared. The treatment or intervention phase typically comes after the baseline phase, where the independent variable (a treatment or intervention) is introduced to the unit for the first time.
3) Measurement stability must be achieved before any phase changes. For example, in a four phase ABA′B′ reversal design the data of the outcome variable must be stable before changing phases from baseline (phase A) to the treatment or intervention phase (phase B). Phase B data must also become stable before changing to the withdrawal of treatment or intervention phase (phase A′), and then reach stability once more before reintroducing the treatment or intervention (phase B′). More specifically, measurement stability refers to a lack of trend (i.e., lack of consistent increases or decreases) in the outcome variable, or a trend that is opposite in direction of the anticipated treatment or intervention effect, and a steady range of variability.
4) Within each phase, there must be a deliberate manipulation of one and only one independent variable at a time.

5) Replications across units, settings, and practitioners may be included to demonstrate evidence of external validity, resulting in SCED studies with multiple units. In other words, SCED studies may include multiple units where the treatment effect on each unit is investigated via a SCED.

There are many variations of SCEDs. A SCED that is designed so that the replication of treatment or intervention effects can be observed systematically has sound internal validity. For example, the treatment effect should be present when the treatment is applied, not present when the treatment is withdrawn, and present again when the treatment is reintroduced, while holding all potential covariates constant. This common SCED is generally called the ABA′B′ or reversal design, and has previously been described. This design has sound internal validity because it rules out many alternative explanations for the outcome change. Specifically, once treatment is withdrawn (phase A′) data is expected to return to what is observed in the baseline (phase A), and similar data from phase B is expected when treatment is reintroduced (phase B′). Data from phases A′ and B′ are expected to be similar to those observed in phase A and B, but not identical due to potential carry over effects. If the treatment or intervention cannot be withdrawn for practical or ethical reasons or the outcome of interest is not reversible (e.g. learning behavior), the effect of the treatment may be replicated across settings, behaviors, or participants, among other possibilities; this situation produces a multiple-baseline design. If multiple baselines are not feasible, a different treatment may be introduced instead of withdrawing the original treatment, which is known as an alternating treatment design. In an alternating treatment design, the effect of the treatment or intervention is generally replicated after baseline and after the newly introduced treatment (i.e. ABCB′ design). More detailed account of SCEDs can be found in Kazdin (2011), Barlow et al. (2008), and Gast and Ledford (2014). In terms of evaluating the proposed effect size statistics for a common SCED, this thesis focused on the ABA′B′ design because it is one of the most popular types of SCEDs and is widely used in psychology.

SCEDs, RCTs, and the EBPP Movement in Psychology

SCEDs. An example of a renewed interest in utilizing SCEDs within psychology can be found in the field of neurorehabilitation which "…is concerned with improving function and enhancing community participation after acquired brain impairment" (Perdices & Tate, 2009, p.904). Clinicians working in neurorehabilitation are often presented with the challenge of treating individuals suffering from a wide variety of deficits and disabilities because individuals with similar brain impairment may manifest deficits and disabilities differently (Perdices & Tate, 2009). Furthermore, clinicians are usually required to have an individualized treatment for each client as treatment strategies depend on the specific deficits and disabilities as well as the individual's personal rehabilitation goals (Perdices & Tate, 2009). Therefore, every individual, regardless of their similarity in brain impairment, will most likely require a variation or an entirely different treatment (Perdices & Tate, 2009). Given the vast dissimilarity in the population served, treatment options, and outcomes, SCEDs are essential for evaluating treatment efficacy in fields such as neurorehabilitation as group-based designs may not be feasible. Accordingly, the utility of SCEDs in treatment or intervention evaluation is increasingly vital as the integration of best available treatment evidence with clinical practice is made a priority by the evidence-based practice in psychology (EBPP) movement (APA Presidential Task Force on Evidence-Based Practice, 2006).

Randomized controlled trials. Traditionally, randomized controlled trials (RCTs) have been the gold standard for providing evidence for various treatment efficacies (e.g. Robb, 2013). Yet RCTs, as with all research methodologies, have their own limitations. RCTs are generally a type of group-based experimental design that estimates the treatment effect from the comparison of different groups of individuals. The result of an RCT is an average treatment effect for different individuals. In the context of fields such as neurorehabilitation, one limitation of RCTs is that they do not directly focus on the heterogeneity within individuals across different treatment conditions. In other words, because not all individuals exposed to the same treatment will respond to the treatment equivalently, the results of RCTs may not generalize to specific individuals. That is, between-person treatment effects may not equal within-person treatment effects. Clinicians are generally not as interested if a treatment works on average, but whether a treatment will work for a specific client (Perdices & Tate, 2009). Hence, the results of RCTs may be of limited practical use since they may not be relevant to clinical practice (Johnston & Smith, 2010). After all, one of the goals of evidence-based practice is to get the best evidence to inform treatment decisions about the individual under care (Johnston & Smith, 2010). Another disadvantage of RCTs is the need for relatively large homogeneous groups (Busse, Kratochwill, & Elliott, 1995; Gingerich**,** 1984). Furthermore, ethical problems may also prevent random assignment of participants to a treatment or control group (Gingerich**,** 1984). So in fields such as neurorehabilitation, where there is a large variation in the population served, treatment options, and outcomes, RCTs may not be feasible. Moreover, not all treatments need to be evaluated via RCTs (e.g. Smith & Pell, 2003). For instance, a RCT is not necessary and probably will not be funded to evaluate the efficacy of using wheelchairs or grab bars by individuals with movement disabilities (Johnston & Smith, 2010). Nevertheless, treatments or interventions (notwithstanding how simple they may appear to be) should be evaluated so that decisions about whether to utilize them can be based on their efficacy. In these situations where group-based designs may not be feasible, SCEDs are a good alternative to RCTs.

Unlike RCTs, SCEDs do not focus on the average treatment effect across individuals but instead explicitly focus on within-individual variability. In other words, SCEDs emphasize change at the unit or individual level by comparing the treatment effects on a single unit or individual over different measurement occasions. SCEDs are also flexible and are able to accommodate research areas with high variability or low prevalence rates commonly found in fields such as clinical psychology, occupation therapy, special education, social work, and medicine, among others. In fact, Guyatt et al. (2000) have put SCEDs that use randomization and blind components above RCTs in their hierarchy of best potential for providing evidence on various treatment efficacies. For example, phases, treatments and practitioners, among others, can be randomly assigned to the unit under investigation. Additionally, the What Works Clearinghouse has recently accepted SCEDs as an adequate research methodology to provide evidence on various treatment and intervention efficacies (Kratochwill et al., 2010). Yet SCEDs also have limitations. The external validity of the findings obtained by SCEDs has been most often critiqued as inferior because a single SCED provides very little, if any, information about between-individual variability. In other words, the finding of a SCED is highly relevant to the unit or individual under investigation, but the finding may not be generalizable to other participants or settings. Therefore, little information is available to draw inferences about the overall efficacy of a treatment or intervention. This disadvantage, however, can be remediated by meta-analyzing results from separate SCEDs (Gingerich**,** 1984; Robey, Schultz, Crawford, & Sinner, 1999).

Research Synthesis, Meta-analysis, and SCEDs

Traditional research synthesis and meta-analysis generally only include group-based experimental designs such as RCTs (Busse et al. 1995). As meta-analytic techniques for SCEDs have only recently been developed, one of the reasons for past exclusions of SCEDs is the lack of established appropriate quantitative methodology (Allison & Gorman, 1993; Center, Skiba, & Casey, 1985). The primary data for a meta-analysis is an effect size (ES) estimate from each study. To date, there is a lack of consensus as to the appropriate measure of ES of SCEDs, which is a preliminary challenge and prerequisite for further advancement (Lane & Carter, 2013; Shadish & Rindskopf, 2007).

There are various drawbacks to deliberately excluding SCEDs from a research synthesis and meta-analysis (Allison & Gorman, 1993; Scruggs & Mastropieri, 2013; Shadish et al. 2008). The most pronounced of these drawbacks is a systematic lack of evidence taken into consideration, especially when a given research topic has used a relatively large amount of SCEDs compared to group-based designs. In such cases, EBPP goals may be best met if evidence of a treatment or intervention efficacy produced by SCEDs and group designs are both considered (Robey et al. 1999) and potentially synthesized collectively so that a more comprehensive assessment of treatment efficacy can be obtained. Examples of various efforts to increase the awareness of SCEDs in meta-analysis can be seen with the publication of special issues in the *Evidence-Based Communication Assessment and Intervention* journal (volume 2, issue 3, 2008), the *Journal of Behavioral Education* (volume 21, issue 3, 2012), the *Remedial and Special Education* journal (volume 34, issue 1, 2013), the *Journal of Applied Sport Psychology* (volume 25, issue 1, 2013), and the *Journal of School Psychology* (volume 52, issue 2, 2014), among others. Alongside these efforts, research syntheses and meta-analyses of solely SCEDs have also begun to appear more frequently (Maggin, O'Keeffe, & Johnson, 2011).

Research synthesis and meta-analysis may be used to summarize and produce evidence of external validity for SCEDs as well as to obtain a more precise estimate of the magnitude of the treatment or intervention effect under investigation. Additionally, by using moderator analysis, the meta-analysis findings can also fine-tune the generalizability of a treatment or intervention (Burns, 2012; Gingerich, 1984). For instance, factors that may moderate the effectiveness of a treatment or intervention such as participant characteristics, settings, or practitioners can be coded and evaluated. Having more specific information as to for whom, where, and when a treatment or intervention works best is part of the essential dimensions of EBPP (APA Presidential Task Force on Evidence-Based Practice, 2006).

SCED data and the proposed Effect Size statistics

The data produced by SCEDs are similar to time-series data: repeated measurements on a single unit are represented as a sequence of data points. The differences, however, lie in the length of the series (series of SCEDs are shorter), the interruption and potential modification of the series during phase changes (after a phase change, the series may change in variability, level, trend, or any combination thereof). Thus, the data produced by SCEDs may be viewed as a special case of time-series and have been referred to as interrupted time-series data (Gingerich, 1984). Consequently, interrupted time-series data may inherit many of the characteristics of a time-series such as the presence of autocorrelation (also known as autoregression, serial dependency, serial correlation, lagged correlation, etc.). Autocorrelation is the degree of correlation between past and future values in a time series observation. Said differently, it is a lack of independence among the observations in a time series. If these correlations are not taken into account, the time series observations may appear more homogeneous. For example, if a child was ill at the beginning of a time series study, the effect of being ill may influence the multiple consecutive observations in the time series study (Ferron, 2002). Depending on the study and outcome measure, this non-independence across observations may result in a positively or negatively autocorrelated error structure. Ignored positive or negative autocorrelation potentially leads to smaller or larger standard deviation, thus leading to an overestimation or underestimation of a standardized ES that is dependent on parametric assumptions.

An ES may be defined as “a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest” (Kelley & Preacher, 2012, p. 4). In light of this broad definition, an ES may convey information in terms of standardized unit of change, variance accounted for, change in mean, change in raw units, and change in variability, among others, depending on the research question at hand. Examples of popular ES are the standardized mean difference ES (e.g. Cohen’s *d*) and the standardized linear association ES (e.g. Pearson’s product-moment correlation, *r*). Many ES statistics have been proposed for SCED data, the present thesis focused on SCED ESs that are designed to convey information in terms of standardized unit of change, mean change, and variance accounted for because finding an ES for SCEDs which is comparable to group-based experimental design ESs is one of the goals of the thesis. The type of ES statistics proposed for SCED data can largely be divided into two major categories: regression-based and non-regression-based, with the latter being more popular in general. Below, various regression- and non-regression-based ESs are reviewed. For ease of presentation, explanation of ES is first based on data produced by a simple two-phase SCED (AB), which is later generalized to a four-phase SCED (ABA′B′).

Regression-based ES: Simple linear regression. A two-phase (AB) interrupted time-series data can be analyzed via a simple linear regression model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t \ ,$$

where Y_t is the participant's observed score on the dependent variable at time *t* where *t* = 1, 2, …, *T*, β_0 is a constant representing the initial level (intercept), X_{1t} is a dummy-coded variable representing each phase (X_{1t} = 0 for phase A, X_{1t} = 1 for phase B), and ε_t is the difference between the participant's observed score and predicted score at time *t*. The simple linear regression model is easily generalized to accommodate a four-phase (ABA′B′) interrupted time-series data by including the appropriate dummy variables:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t \ , \qquad (1)$$

where Y_t is the participant's observed score on the dependent variable at time *t*, β_0 is a constant representing the initial level (intercept), X_{1t} is a dummy-coded variable representing phase B, X_{2t} is a dummy-coded variable representing phase A′, X_{3t} is a dummy-coded variable representing phase B′ and ε_t is the difference between the participant's observed score and predicted score at time *t*. This equation is now a multiple linear regression model.

Two common types of ES statistics may be obtained from the simple linear regression and multiple linear regression models:

1) The regression coefficient β_1 (i.e. the effect of phase B) in the simple linear regression, and the average of both partial regression coefficients β_1 and β_3 (i.e. the average partial effect of phase B and B′) in the multiple linear regression model shown in Equation 1.

2) The variance accounted for by the simple linear regression model R^2, and the unique variance accounted for by phases B and B′ via squared semi-partial correlation sR^2 obtained from the multiple linear regression model.

The main shortcoming of these ES statistics is their inability to account for the effects of a time-based trend since the time variable is not included in either regression model. If a trend is present, depending on its direction the regression coefficients and squared semi-partial correlation may underestimate or overestimate the magnitude of the actual effect. The simple and multiple linear regression model also assumes homogeneity of variance and an independent error structure. Given the plausible presence of autocorrelation in SCED data, the assumption of independence for these models and all subsequent regression-based models is likely violated.

Regression-based ES: Gorsuch's method. As an improvement to the simple linear regression model above, Gorsuch (1983) proposed a model which accounts for a linear time-based trend for a two-phase (AB) interrupted time-series data:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \varepsilon_t \ ,$$

where Y_t is the participant's observed score on the dependent variable at time *t* where *t* = 1, 2, …, T, β_0 is a constant representing the initial level (intercept), *t* is the time variable (thus β_1 represents the slope of the trend for both phases), X_{1t} is a dummy-coded variable representing phase (X_{1t} = 0 for phase A, X_{1t} = 1 for phase B), and ε_t is the difference between the participant's observed score and predicted score at time *t*. As with the simple linear regression model, this model is also easily generalized to accommodate a four-phase (ABA′B′) interrupted time-series data by including the appropriate dummy variables:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \beta_3 X_{2t} + \beta_4 X_{3t} + \varepsilon_t \ , \qquad (2)$$

where Y_t is the participant's observed score on the dependent variable at time *t*, β_0 is a constant representing the initial level (intercept), *t* is the time variable, X_{1t} is a dummy-coded variable representing phase B, X_{2t} is a dummy-coded variable representing phase A′, X_{3t} is a dummy-coded variable representing phase B′, and ε_t is the difference between the participant's observed score and predicted score at time *t*. Time *t* at B and B′ phases is centered to aid interpretation of the partial regression coefficients. In other words, the mean of t_B time points is subtracted from all time points in phase t_B while the mean of $t_{B'}$ time points is subtracted from all time points in phase $t_{B'}$. For example, if t_A has five time points, 1, 2, 3, 4, and 5, and t_B has five time points, 6, 7, 8, 9, and 10, t_B is centered by subtracting 8 from 6, 7, 8, 9, and 10. Similarly, if $t_{B'}$ has time points, 16, 17, 18, 19, and 20, $t_{B'}$ is centered by subtracting 18 from 16, 17, 18, 19, and 20.

As before, two common types of ES statistics may be obtained from Gorsuch's model and the extended model shown in Equation 2:

1) The partial regression coefficient β_2 (i.e. the partial effect of phase B) in Gorsuch's model, and the average of both partial regression coefficients β_2 and β_4 (i.e. the average partial effect of phase B and B′) in the extended model shown in Equation 2.

2) The semi-partial correlations, sR^2, via Gorsuch's model (i.e. the unique variance accounted for by phase B) and the extended model shown in Equation 2 (i.e. the unique variance accounted for by phases B and B′).

The main shortcoming of these ES statistics is their inability to account for the effects of a different slope for *t* in each phase. If there is a different slope in each phase, the ES statistics may underestimate or overestimate the actual treatment effect.

Regression-based ES: Center, Skiba, & Casey's method. As a further improvement, Center, Skiba, and Casey (1985) suggested the use of a piecewise regression model for a two-phase (AB) interrupted time-series data:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \beta_3 X_{1t}(t_B) + \varepsilon_t \ ,$$

where Y_t is the participant's observed score on the dependent variable at time *t*, β_0 is a constant representing the initial level (intercept), *t* is the time variable, X_{1t} is a dummy-coded variable representing each phase (X_{1t} = 0 for phase A, X_{1t} = 1 for phase B), $X_{1t}(t_B)$ is an interaction term which allows for a different slope for the time-based trend in phase B (thus β_3 represents the slope in phase B), t_B is a subset of the time variable for data points in phase B only, and ε_t is the difference between the participant's observed score and predicted score at time t. As before, this model is also easily generalized to accommodate a four-phase (ABA′B′) interrupted time-series data by including the appropriate dummy and interaction variables:

$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \beta_3 X_{2t} + \beta_4 X_{3t} + \beta_5 X_{1t}(t_B) + \beta_6 X_{2t}(t_{A'}) + \beta_7 X_{3t}(t_{B'}) + \varepsilon_t$, (3)

where Y_t is the participant's observed score on the dependent variable at time t, β_0 is a constant representing the initial level (intercept), t is the time variable, X_{1t} is a dummy-coded variable representing phase B, X_{2t} is a dummy-coded variable representing phase A′, X_{3t} is a dummy-coded variable representing phase B′, $X_{1t}(t_B)$ is an interaction term which allows for a different slope in phase B, $X_{2t}(t_{A'})$ is an interaction term which allows for a different slope in phase A′, $X_{3t}(t_{B'})$ is an interaction term which allows for a different slope in phase B′, and ε_t is the difference between the participant's observed score and predicted score at time t. As with the extension to Gorsuch's method, time t at B and B′ phases are also centered to aid interpretation of the partial regression coefficients.

With the presence of interaction terms in Center et al.'s (1985) model and the extended model in equation 3, only the first-order regression coefficients are considered further: the partial regression coefficient β_2 represents the partial effect of phase B in Center et al.'s (1985) model, while the average of β_2 and β_4 represents the average partial effect of phase B and B′ in the extended model shown in Equation 3. A common criticism of this method lies in the multiple choices of partial regression coefficients available to represent the ES estimate. Yet, the different partial regression coefficients represent different aspects of treatment effects. Specifically, with Center et al.'s model ($Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \beta_3 X_{1t}(t_B) + \varepsilon_t$) as an example, β_3 may be regarded as an ES for the magnitude of the slope change in the intervention phase, and not the instantaneous magnitude of level change in treatment effect, which is given by β_2. In terms of the extended model shown in Equation 3, β_2 and β_4 are the instantaneous magnitude of treatment effect for phase B and B′ respectively with β_5, β_6, and β_7 are the time-trend slope changes in phase B, A′, and B′ respectively.

Regression-based ES: Allison & Gorman's method. Allison and Gorman (1993) proposed a refinement to the method of Center et al. (1985) for a two-phase (AB) interrupted time-series data. Allison and Gorman's method involves multiple steps designed to remove the effects of a time-based trend before estimating an ES in terms of R^2. Doing so may be especially beneficial if a trend is in the same direction of the slope for phase B which can cause an overestimation of the ES. The time-based trend is estimated from the baseline data only, and is removed from the analysis of subsequent phases as described below:

1) Fit a simple linear regression model to the baseline phase:

$$Y_{t_A} = \beta_0 + \beta_1 t_A + \varepsilon_{t_A} \ ,$$

2) Calculate the residuals from the baseline phase:

$$\hat{\varepsilon}_{t_A} = Y_{t_A} - \hat{Y}_{t_A} \ ,$$

3) Generate predicted values in the treatment phase from the time-based trend estimate of step 1. In other words, the value of the regression coefficient β_1 from step 1 is used as the value for the regression coefficient β_1 here:

$$\hat{Y}_{t_B} = \hat{\beta}_0 + \hat{\beta}_1 X_{1t}(t_B) \ ,$$

4) Calculate the residuals from the treatment phase:

$$\hat{\varepsilon}_{t_B} = Y_{t_B} - \hat{Y}_{t_B} \ ,$$

These residuals or "detrended data" obtained in steps 2 and 4 are proposed to deal with the problem of potential overestimation due to the presence of a time-based trend and a slope that is in the same direction. The detrended data are used in the subsequent steps:

5) Calculate the zero-order correlation between X_{1t} and the residuals of baseline phase ($\hat{\varepsilon}_{t_A}$) and the zero-order correlation between $X_{1t}(t_B)$ and the residuals of treatment phase ($\hat{\varepsilon}_{t_B}$) and compare their signs. If the signs are in the same direction, proceed to step 6. If the signs are not in the same direction, skip to step 10.
6) Fit a multiple linear regression model to the residuals of baseline and treatment phase:

$$\hat{\varepsilon}_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1t}(t_B) + \varepsilon_t \ ,$$

   Where $\hat{\varepsilon}_t$ is an element from the vector that concatenates $\hat{\varepsilon}_{t_A}$ for the baseline phase and $\hat{\varepsilon}_{t_B}$ for the intervention phase.
7) Obtain the R^2 statistic from the multiple linear regression model in step 6.
8) If preferred, Allison and Gorman suggest that the R^2 statistic may be converted into an *F*-ratio, which may then be converted to a *d* or *r* ES statistic.
9) The same sign obtained in step 5 may be assigned to the *d* or *r* statistic, resulting in an ES that represents the magnitude and direction of the effect.
10) If the signs of the zero order correlations are not in the same direction, Allison and Gorman (1993) proposed to simply look at the effect of the level as recommended by Center et al. (1985). The rationale is that overestimating due to a time-based trend should not be a problem if the slope in phase B is in an opposite direction.

A potential problem with this method may be shrinkage. De-trended data may produce residuals without the appropriate variability, thus leading to negatively biased standard errors. This method may accommodate a four-phase (ABA′B′) interrupted time-series data by repeating the appropriate steps mentioned previously for different sets of (AB) phases. For example, the R^2 estimate may be obtained for AB, A′B′, and AB′ phases separately.

Non-regression-based ES: Percentage of non-overlapping data (PND). The PND proposed by Scruggs, Mastropieri, and Casto (1987) is one of the earliest ES statistics designed for SCEDs. Here, the calculation of PND is described for a two-phase (AB) interrupted time-series data:

1) If an increase in outcome measurement is expected, the PND is the percentage of data points in the treatment phase (phase B) that surpass the highest data point in the baseline phase (phase A).

2) If a decrease in outcome measurement is expected, the PND is the percentage of data points in the treatment phase (phase B) that are lower than the lowest data point in the baseline phase (phase A).

Similar to Allison and Gorman's (1993) method, the PND statistic may also accommodate four-phase (ABA′B′) interrupted time-series data by separate calculations for different sets of (AB) phases. For example, PND may be calculated for AB, A′B′, and AB′ phases separately. Additionally, an average of PNDs for both AB and A′B′ is also possible.

The main limitations of PND are the statistic's sensitivity to outliers and insensitivity to the potential effects of a time-based trend and differing time-trend slopes in different phases of the data. Although many have critiqued the statistical properties of PND, it remains the most common ES statistic used in the meta-analyses of SCEDs (Beretvas & Chung, 2008).

Nonregression-based ES: Standardized mean difference (SMD) methods. Another one of the earliest used ES for a two-phase (AB) interrupted time-series data is in a form that closely resembles Cohen's *d*:

$$\delta_{\mathrm{SMD1}} = \frac{\bar{Y}_B - \bar{Y}_A}{SD_A}, \qquad (4)$$

where

$$SD_A = \sqrt{\frac{t_A \sum Y_t^2 - (\sum Y_t)^2}{t_A(t_A - 1)}},$$

and Y is the participant's observed score on the dependent variable, $\bar{Y}_A$ is the average of participant's observed score on the dependent variable at phase A, $\bar{Y}_B$ is the average of participant's observed score on the dependent variable at phase B, SD_A is the standard deviation of phase A, and t_A is the subset of the time variable for data points in phase A only. The method works by taking the difference between phase means and dividing it by the standard deviation of the baseline phase. This way, no assumptions are made regarding the distributional form of the observations and homogeneity of variance across phases. Busk and Serlin (1992) proposed standardizing the mean differences with the variability of the entire data:

$$\delta_{SMD2} = \frac{\bar{Y}_B - \bar{Y}_A}{SD_{pooled}}, \quad (5)$$

Where

$$SD_{pooled} = \sqrt{\frac{t \sum y_t^2 - (\sum Y_t)^2}{t(t-1)}},$$

and t is the time variable representing the total number of data points in the interrupted time series. This way, parametric assumptions such homogeneity of variance across both phases are in effect, and if desired, a confidence interval may be constructed. As with Allison and Gorman (1993) and PND statistics, both δ_{SMD1} and δ_{SMD2} may accommodate a four-phase (ABA′B′) interrupted time-series data by separate calculations for different sets of (AB) phases. For example, δ_{SMD1} and δ_{SMD2} may be calculated for AB, A′B′, and AB′ phases separately, in addition to an average of δ_{SMD1} and δ_{SMD2} statistics for the AB and A′B′ phases.

Similar to the PND, the main shortcoming of both δ_{SMD1} and δ_{SMD2} is their inability to account for potential effects of a time-based trend and differing slopes in different phases of the data. Furthermore, parametric assumptions may be violated for δ_{SMD2}. Nevertheless, δ_{SMD} type of

ES is the second-most common ES used for a meta-analysis of SCEDs (Beretvas & Chung, 2008).

Non-regression-based ES: standardized mean difference. Hedges, Pustejovsky, and Shadish (2012) proposed an ES (δ_{HPS}) that is comparable to Cohen's *d* (i.e. scaled to the same metric and utilizing equivalent parameter estimates in its calculation). A minimum of three SCED units are needed for estimating the variance of δ_{HPS}. Therefore, this ES may not be utilized with SCED studies with fewer than three units. Hedges et al. (2012) posited a general framework that encompasses both group-based experimental designs and SCEDs with replication across individuals. Using the same notation as Hedges et al. (2012), the data for the model can be represented as:

$$\begin{array}{lccl} & \text{Phase A} & \text{Phase B} & \\ \text{Control group} & \begin{bmatrix} Y_{11} & \cdots & Y_{1n} \\ \vdots & & \vdots \\ Y_{m1} & \cdots & Y_{mn} \end{bmatrix} & \begin{bmatrix} Y_{1(n+1)} & \cdots & Y_{1(2n)} \\ \vdots & & \vdots \\ Y_{m(n+1)} & \cdots & Y_{m(2n)} \end{bmatrix} & \text{treatment group} \end{array}$$

Here, Y_{ij} is the j^{th} observation from the i^{th} individual, where $j = 1, \ldots, n$ represents the baseline or control phase, and $j = n + 1, \ldots, 2n$ represents the treatment phase. The sample size $i = 1, \ldots, m$ is 1 for SCEDs and *m* for group-based experimental designs. From that general framework, an ES comparable to Cohen's *d* is derived:

$$ES = \frac{\overline{D}}{S},$$

where $\overline{D}$ is the average difference between the treatment (B) and baseline (A) phase that is averaged across all units:

$$\overline{D} = \frac{1}{m}\sum_{i=1}^{m} \left(\frac{1}{n}\sum_{t=n+1}^{2n} Y_{it} - \frac{1}{n}\sum_{t=1}^{n} Y_{it} \right),$$

and S^2 is the average of the sum of squared pooled deviation scores both phases across all units:

$$S^2 = \frac{1}{2n(m-1)} \sum_{t=1}^{2n} \sum_{i=1}^{m} (Y_{it} - \bar{Y}_{.t})^2 \,,$$

$$\bar{Y}_{.t} = \frac{1}{m} \sum_{i=1}^{m} Y_{it}$$

This ES is generalizable to a four-phase (ABA′B′) design by taking the average of the difference between phases AB and A′B′ and dividing that by the average sum squared pooled deviation scores across all four phases. δ_{HPS} however, operates under the assumptions that Y_{ij} is normally distributed, that no time-based trend is present, and only weak first-order autocorrelation may be present in the data.

In sum, SCEDs are a set of experimental designs that produce data with similar characteristics to a time-series. SCED data is also known as interrupted time-series, and, as a special case of time-series, it inherits characteristics such as the presence of autocorrelation across non-independent observations. Due to the challenging nature of SCED data, there is a lack of consensus as to the optimal ESs for application in meta-analyses. Because ESs are the primary data for a meta-analysis, this is a preliminary challenge and requirement for further advancement. Thus, the present thesis evaluated the statistical properties of various ES statistics for a four phase ABA′B′ design in terms of their ability to estimate and represent a treatment effect across a variety of simulated short interrupted time-series data. More specifically, the ESs were computed from simulated data types that exhibited autocorrelation, a time-based trend, and different slopes for time-trends within the phases, and evaluated visually via graphs. Finally, the possibility of an amalgamation of both SCEDs and group designs in a meta-analysis was also discussed.

Method

Simulation Procedure

The ESs were evaluated via a simulation study. The simulation procedure was based on Manolov and Solanas (2008) but extended to accommodate a four phase ABA′B′ design. Firstly, interrupted time-series data representative of a four phase ABA′B′ reversal design were simulated 10,000 times for each cell of the simulation design for data types consisting of no treatment effects followed by data types with non-zero treatment effects. Within all data types, there were 19 degrees of autocorrelation ranging from -0.9 to 0.9 (including a condition with no autocorrelation). In other words, the simulation had a 5 x 3 x 19 design where the first factor investigated was the type of effect (5), while the second factor was phase length (3) and the third factor was the degree of autocorrelation (19). Secondly, the ES statistics and their raw bias were calculated and averaged over replications for each cell. Finally, to visualize the effects of autocorrelation and data type on the ESs, the ESs were plotted against degree of autocorrelation. The 26 ES statistics considered in the simulation were:

1) the average of both partial regression coefficients corresponding to a level change in phases B and B′ from Equation 1 ($\bar{\beta}_1$),
2) the average of the standardized regression coefficients of $\bar{\beta}_1$ ($\bar{\beta}_1^*$),
3) the unique variance accounted for by phases B and B′ from Equation 1 (sR_1^2),
4) the average of both partial regression coefficients corresponding to a level change in phases B and B′ from Equation 2 ($\bar{\beta}_2$),
5) standardized regression coefficient of $\bar{\beta}_2$ ($\bar{\beta}_2^*$),
6) the unique variance accounted for by phases B and B′ from Equation 2 (sR_2^2),

7) the average of both partial regression coefficients corresponding to a level change in phases B and B′ from Equation 3 ($\bar{\beta}_3$),
8) standardized regression coefficient of $\bar{\beta}_3$ ($\bar{\beta}_3^*$),
9) Allison and Gorman's R^2 for the first set of AB phases (AG-R^2-AB),
10) Allison and Gorman's R^2 for the second set of A′B′ phases (AG-R^2-A′B′),
11) Allison and Gorman's R^2 for the first A and last B′ phases (AG-R^2-AB′),
12) the average of Allison and Gorman's R^2 for AB and A′B′ phases ($\overline{\text{AG} - R2}$),
13) PND for the first set of AB phases (PND-AB),
14) PND for the second set of A′B′ phases (PND-A′B′),
15) PND for the first A and last B′ phases (PND-AB′),
16) the average of PNDs for AB and A′B′ phases ($\overline{PND}_1$),
17) the average of PNDs for AB and AB′ phases ($\overline{PND}_2$),
18) unpooled SMD from Equation 4 for the first set of AB phases (δ-AB),
19) unpooled SMD from Equation 4 for the second set of A′B′ phases (δ-A′B′),
20) unpooled SMD from Equation 4 for the first A and last B′ phases (δ-AB′),
21) the average unpooled SMDs from Equation 4 for AB and A′B′ phases ($\bar{\delta}$),
22) pooled SMD from Equation 5 for the first set of AB phases (δ_p-AB),
23) pooled SMD from Equation 5 for the second set of A′B′ phases (δ_p-A′B′),
24) pooled SMD from Equation 5 for the first A and last B′ phases (δ_p-AB′),
25) the average pooled SMDs from Equation 5 for AB and A′B′ phases ($\bar{\delta}_p$),
26) Hedges, Pustejovsky, and Shadish's standardized mean difference (δ_{HPS})

Detailed information regarding the data generation mechanism and data types is given below.

Data Generation

Interrupted time-series data were generated according to a variation of the standard multiple linear regression model which allowed for a time-based trend, differing time-trend slopes for each phase, as well as autocorrelation in the errors:

$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{1t} + \beta_3 X_{2t} + \beta_4 X_{3t} + \beta_5 X_{1t}(t_B) + \beta_6 X_{2t}(t_{A'}) + \beta_7 X_{3t}(t_{B'}) + \varepsilon_t$, (6)

where

$$\varepsilon_t = \varphi * \varepsilon_{t-1} + u_t ,$$

and Y_t is the participant's observed score on the dependent variable at time t, β_0 is a constant representing the initial level at phase A (intercept), t is the time variable, X_{1t} is a dummy-coded variable representing phase B, X_{2t} is a dummy-coded variable representing phase A′, X_{3t} is a dummy-coded variable representing phase B′, $X_{1t}(t_B)$ is an interaction term which allows for a different slope in phase B, $X_{2t}(t_{A'})$ is an interaction term which allows for a different slope in phase A′, $X_{3t}(t_{B'})$ is an interaction term which allows for a different slope in phase B′, and ε_t is the difference between the participant's observed score and predicted score at time t. The error term ε_t encompasses φ representing the degree of autocorrelation, ε_{t-1} a first-order autoregressive effect, and $u_t \sim N(0,1)$ is the amount of random variation or white noise at time t.

10,000 replications per data type and per degree of autocorrelation were generated using the R statistical program. In other words, a single series was generated and replicated 10,000 times for all ESs per data type and degree of autocorrelation except for δ_{HPS} where 3 series were generated and replicated 10,000 times. A parallel simulation framework[1] was used as the programming environment within R for data generation, ES statistics calculation, and analysis. To avoid dependence between each successive time series, the first 1000 time points were discarded from each time series. In other words, if the actual time points per time series were 20

[1]Designed by Phil Chalmers

($t_A = t_B = t_{A'} = t_{B'} = 5$), a total of 1020 time points were generated per series and the first 1000 discarded. The simulated degrees of autocorrelation, φ, ranged from -0.9 to 0.9 in increments of 0.1 with $\varphi = 0$ denoting the absence of autocorrelation. The R function 'rnorm' was used to generate the white noise (u_t) with a starting seed of 15551.

Data type. 15 types of data were simulated from Equation 6 and within each data type 19 series with degrees of autocorrelation ranging from -0.9 to 0.9 were simulated. A combination of five types of effects (no effects, level change, level and time-based trend, level and time-trend slope within phases, in addition to level, time-based trend and time-trend slope within phases) and three different phase lengths ($t_A = t_B = t_{A'} = t_{B'} = 5,\ t_A = t_B = t = t_{B'} = 10, t_A = t_{A'} = 10\ and\ t_B = t_{B'} = 5$), made up the 15 data types. The population treatment effect of mean difference was 0.3. The effect was generated by setting the partial regression coefficient for the B and B′ phases to 0.3 (Figure 2). The partial regression coefficients for a time-based trend and different slopes within phases were devised so that each had an additional effect on top of the level change (i.e. mean difference). Therefore, when a time-based trend and a different time-trend slope within phases were present in addition to the level change, the total population effect in that phase was greater than 0.3. More specifically, the effect of the time-based trend was $0.3/t/4$) where 4 is the number of phases, and the effect of slope within phases B and B′ was $0.3/((t_B - 1)/2)$ and $0.3/((t_{B'} - 1)/2)$. The time-based trend was designed to represent a 0.3 increase in the difference of the average effect in each pair of AB and A′B′ phase (Appendix A). The time-trend slope within each phase was designed to have a median value of 0.3 so that the difference of the average effect in each pair of AB and A′B′ phase would be 0.3 (Appendix B). Therefore, if a level change and time-based trend were present within phase B, the total effect of phase B is 0.3 + 0.3 = 0.6. Alternatively, if all three effects (level change, time-based trend, and different slope) were present within phase B, the total effect of phase B is 0.3 + 0.3 + 0.3 = 0.9. Table 1 shows the population effects to be estimated by different ES metrics. The intercept was set to zero in all data types as it did not affect the calculation of ESs, and the other partial regression coefficients were varied to produce data types of differing effects. For example, the population data generation models for data types of different effects for a four-phase ABA′B′ with $t_A = t_B = t_{A'} = t_{B'} = 5$ were:

1) No effects data type (see Figure 1):

$$\hat{Y}_t = 0t + 0X_1 + 0X_2 + 0X_3 + 0X_1(t_B) + 0X_2(t_{A'}) + 0X_3(t_{B'})$$

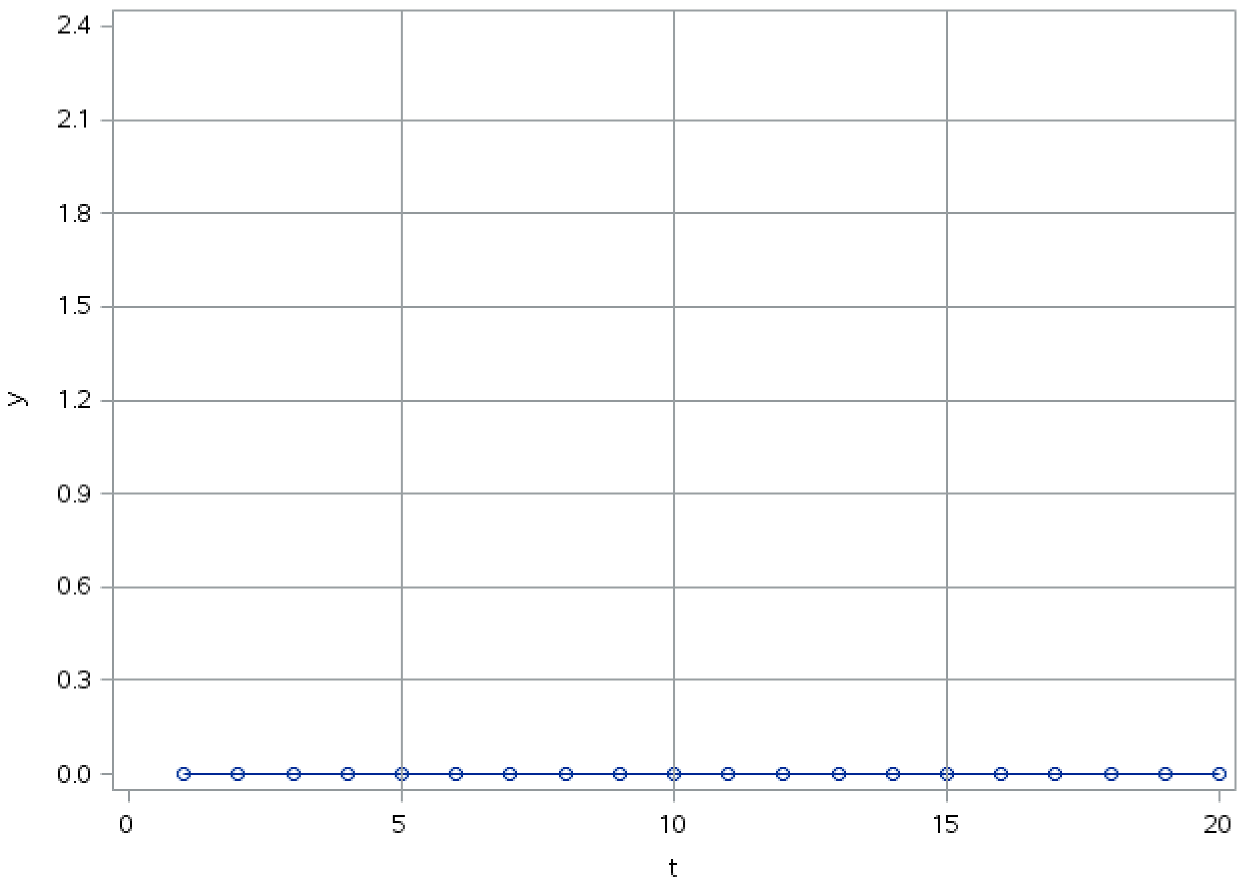


Figure1. No population effect in phases AB and A′B′.

2) Level change data type (Figure 2):

$$\hat{Y}_t = 0t + 0.3X_1 + 0X_2 + 0.3X_3 + 0X_1(t_B) + 0X_2(t_{A'}) + 0X_3(t_{B'})$$

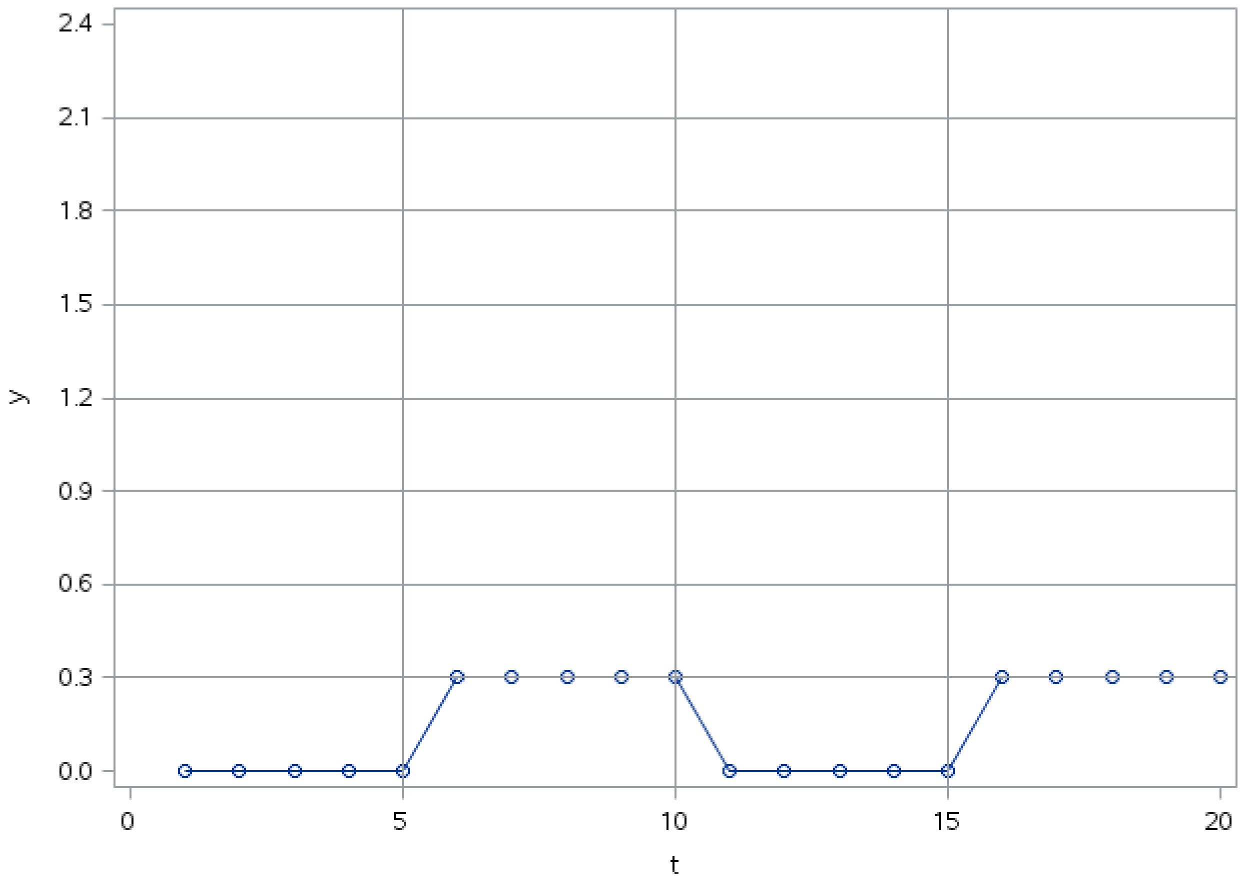


Figure2. Population effect in phases AB and A′B′ is 0.3.

3) Level change and a time-based trend data type (Figure 3):

$$\hat{Y}_t = 0.06t + 0.3X_1 + 0X_2 + 0.3X_3 + 0X_1(t_B) + 0X_2(t_{A'}) + 0X_3(t_{B'})$$

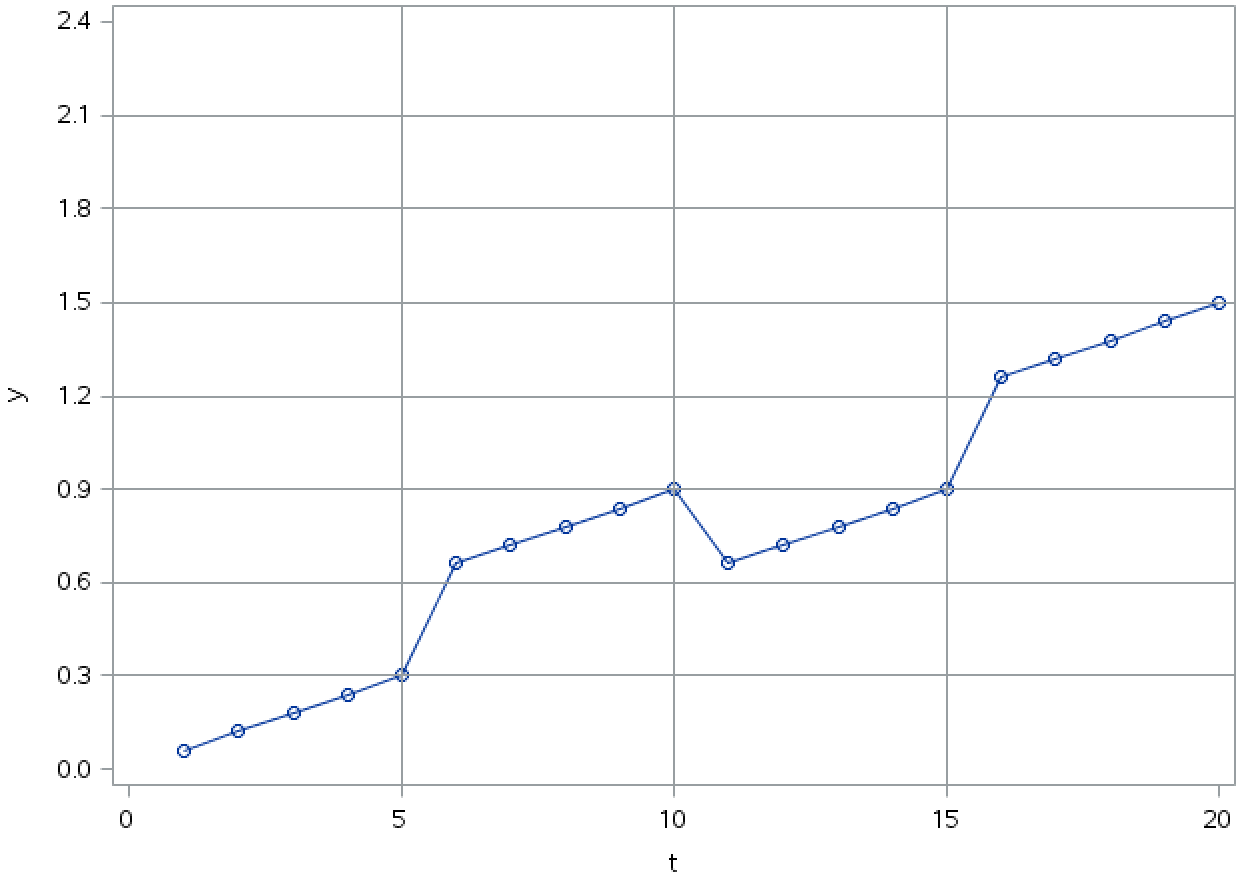


Figure3. Population effect in phases AB and A′B′ is 0.6.

4) Level change and a different slope in B and B′ phases data type (Figure 4):

$$\hat{Y}_t = 0t + 0.3X_1 + 0X_2 + 0.3X_3 + 0.15X_1(t_B) + 0X_2(t_{A'}) + 0.15X_3(t_{B'})$$

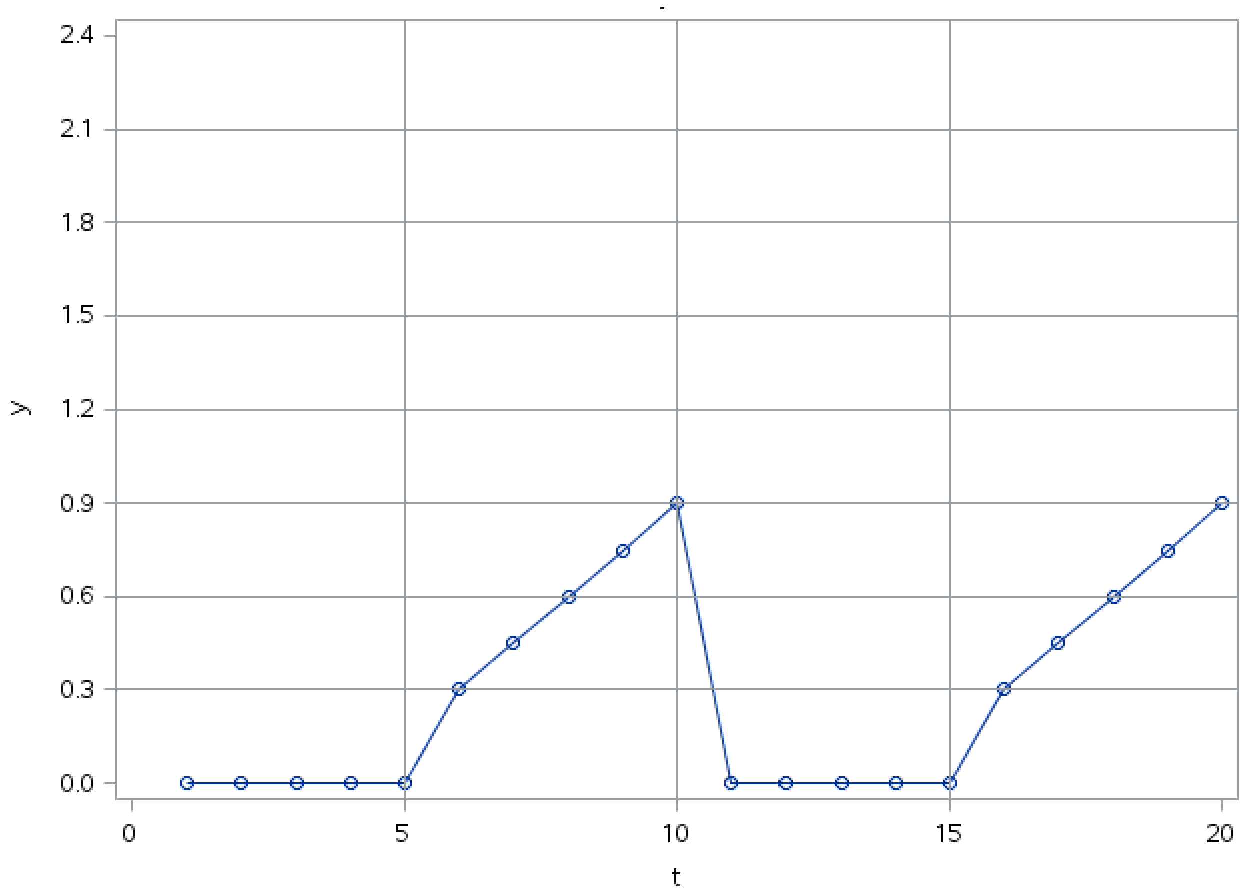


Figure4. Population effect in phases AB and A′B′ is 0.6.

5) Level change, a time-based trend, and a different slope in B and B′ phases data type (Figure 5):

$$\hat{Y}_t = 0.06t + 0.3X_1 + 0X_2 + 0.3X_3 + 0.15X_1(t_B) + 0X_2(t_{A'}) + 0.15X_3(t_{B'})$$

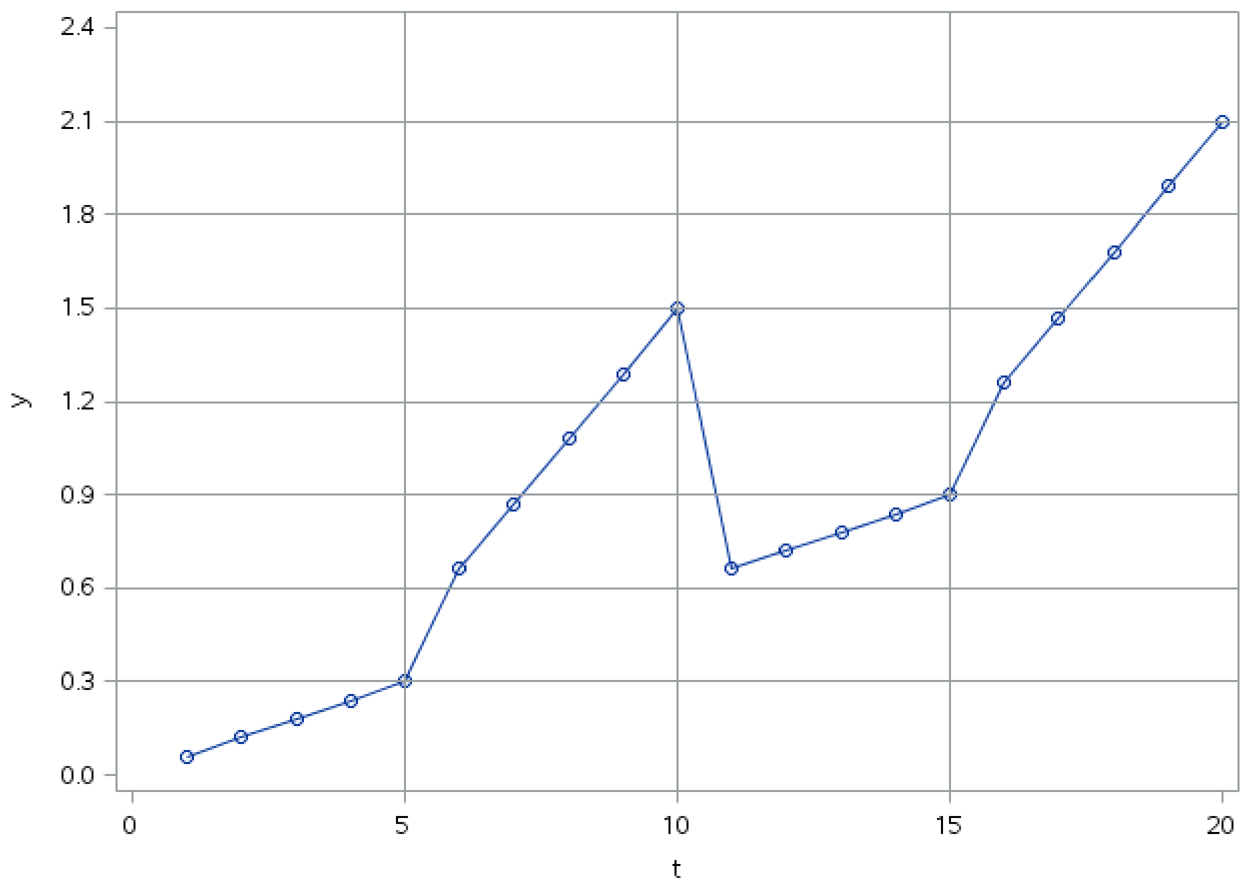


Figure5. Population effect in phases AB and A′B′ is 0.9.

The population effects generated for a four-phase ABA′B′ with $t_A = t_B = t_{A'} = t_{B'} = 10$, and $t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$, were the same as shown above.

Table 1

True magnitude of effect to be estimated by the effect size metrics

	$\bar{\beta}$	$\bar{\beta}*$	δ	R^2
No effects	0.00	0.00	0.00	0.00
Level change only	0.30	0.13	0.30	0.02
Level and time-based trend	0.60	0.26	0.60	0.07
Level and interaction	0.60	0.26	0.60	0.07
Level, time-based trend and interaction	0.90	0.40	0.90	0.16

Note. The interaction represents the differing time-trend slopes in phases B and B′.

Results

For each simulated interrupted time-series, the ES statistics in addition to their raw bias were calculated and averaged over replications. To convey the major results of the simulation, each ES statistic was plotted against degree of autocorrelation ranging from -0.9 to 0.9, and similar graphs were produced for each data type including different phase lengths and effects type. An ideal ES should not be affected by autocorrelation and should be sensitive enough to accurately estimate an effect in short phase lengths.

No Effects Data Type

When no population treatment effect was present, the partial regression coefficients, SMDs and δ_{HPS} ESs were relatively unbiased regardless of the degree of autocorrelation (Figure 6a, Figure 6b, Figure 7). In contrast, the proportion-of-variance ES statistics (R^2 and sR^2) were consistently positively biased, with the degree of bias consistently increasing as autocorrelation increased from -.9 to +.9 (Figure 8, Figure 9). The PND statistics were also consistently positively biased, with more extreme positive and negative autocorrelation resulting in a larger bias resembling a 'U' shape (Figure 10).

Although both partial regression coefficients and SMDs were unbiased, the partial regression coefficients were more variable than SMDs. SMDs may be converted to R^2, and these R^2 metric ESs were again positively biased and affected by autocorrelation such that the bias increased as autocorrelation increased (Figure 11). Yet, compared to Allison and Gorman's R^2, the R^2 converted from SMDs had smaller bias.

Phase length did not extensively augment the relationship of ES statistics with autocorrelation for partial regression coefficients, SMDs, sR^2s, and PNDs. All these ESs exhibited similar relationships with autocorrelation across different phase lengths but with slightly differing variability (Appendix C and Appendix D). However, Allison and Gorman's (1993) R^2 statistics were affected by unequal phase lengths ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$), resulting in less variability among all four calculation methods (Appendix C). When phase lengths were equal, the effects of autocorrelation on Allison and Gorman's R^2 statistics were similar.

Overall, standardized partial regression coefficients and SMDs performed best in terms of being unbiased and unaffected by autocorrelation. In terms of their bias (Figure 12), all values were relatively small but the raw bias for the average of both partial regression coefficients corresponding to a level change in phases B and B′ from Equation 3 ($\bar{\beta}_3$) seemed to be largest among all.

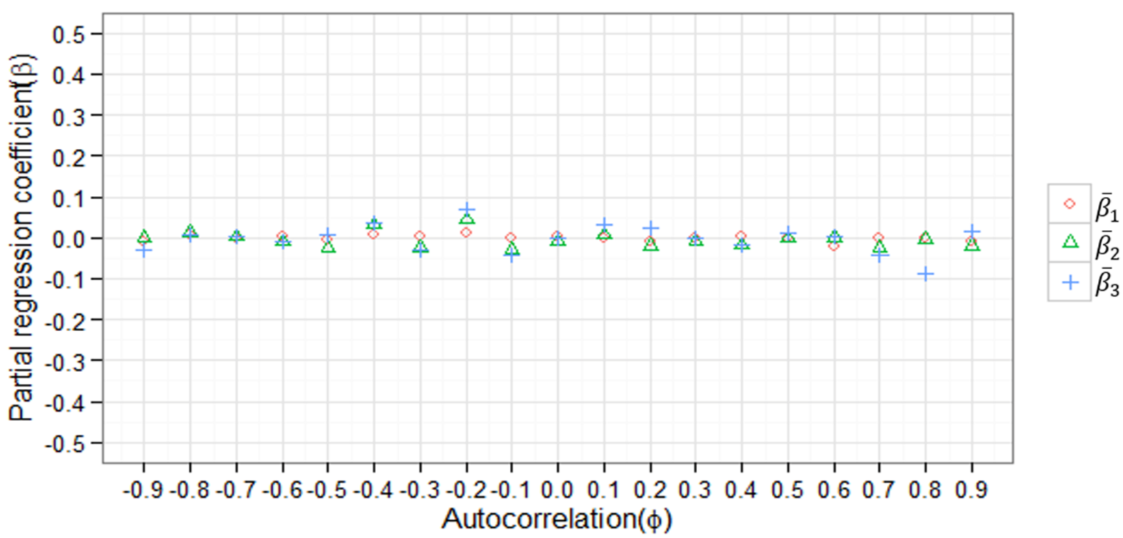


Figure 6a. The effects of autocorrelation on unstandardized partial regression coefficients ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

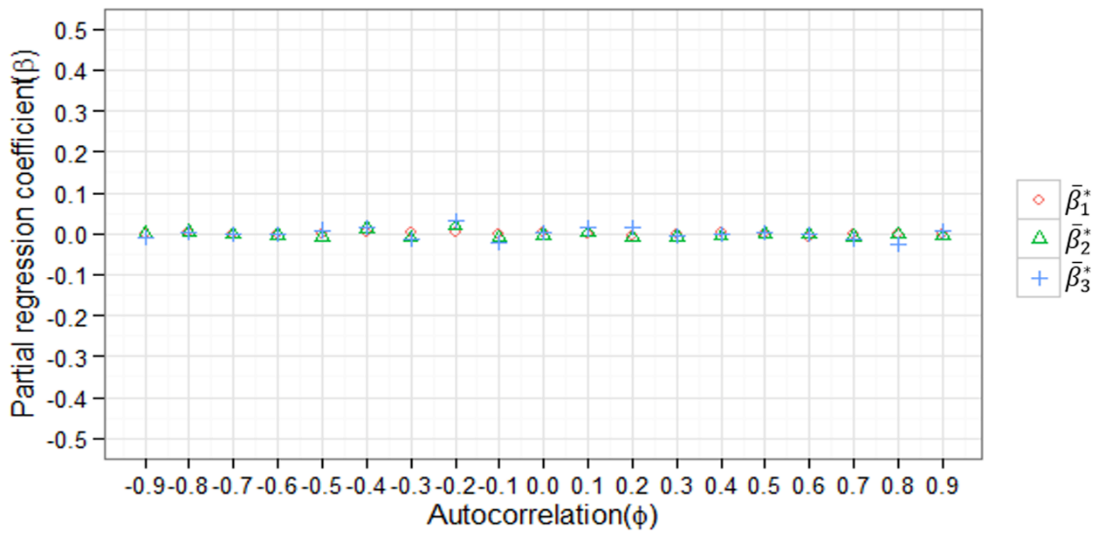


Figure 6b. The effects of autocorrelation on standardized partial regression coefficients ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

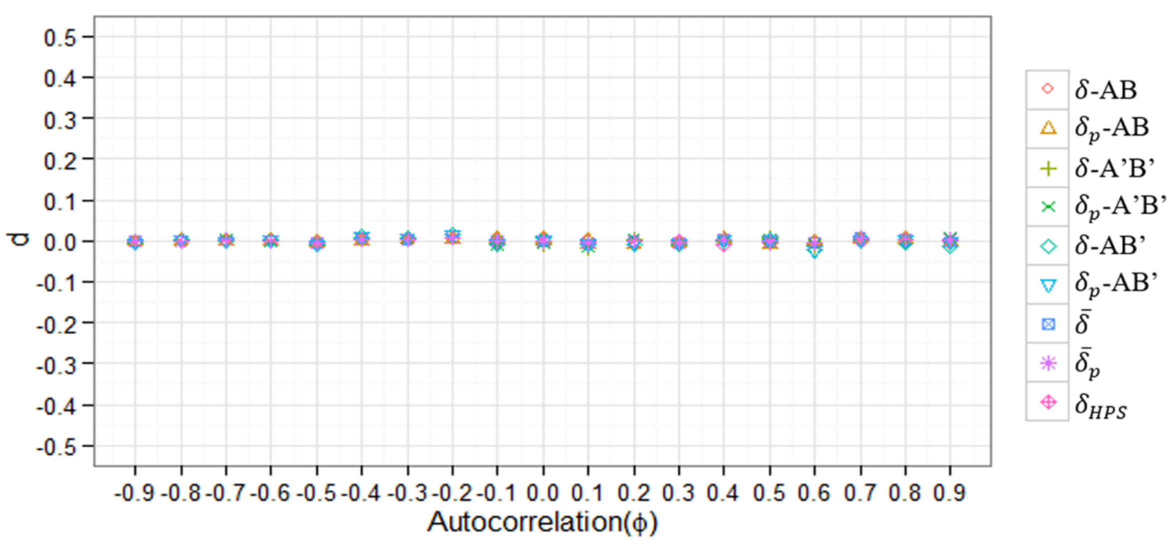


Figure 7. The effects of autocorrelation on standardized mean difference ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

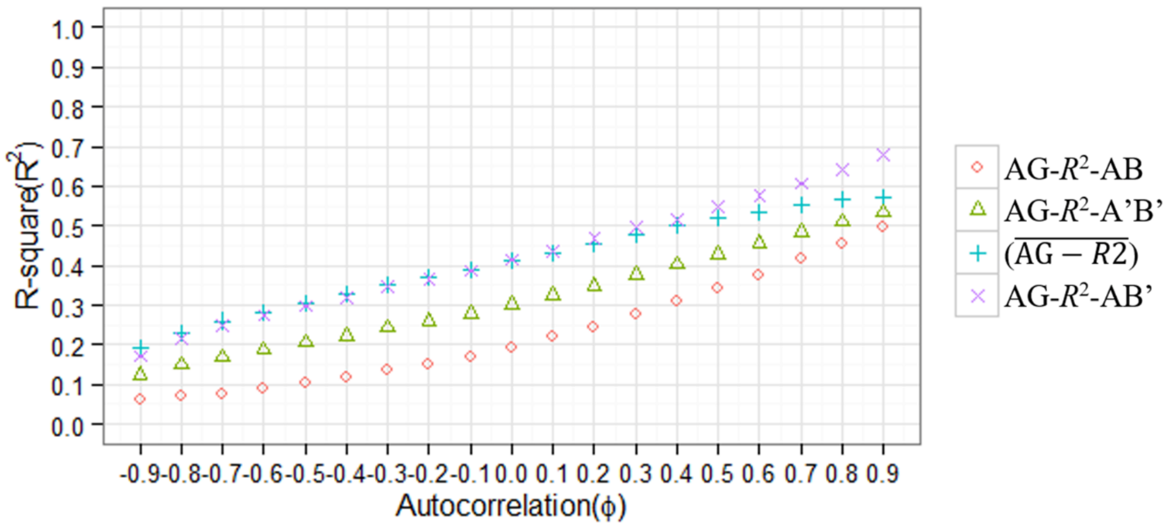


Figure 8. The effects of autocorrelation on Allison and Gorman's R^2 ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

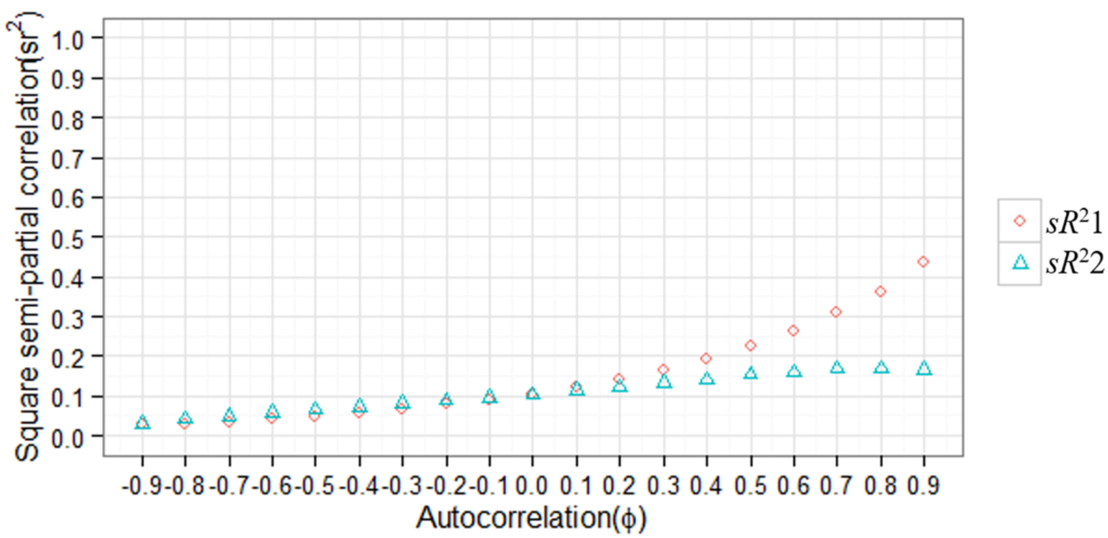


Figure 9. The effects of autocorrelation on squared semi-partial correlation ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

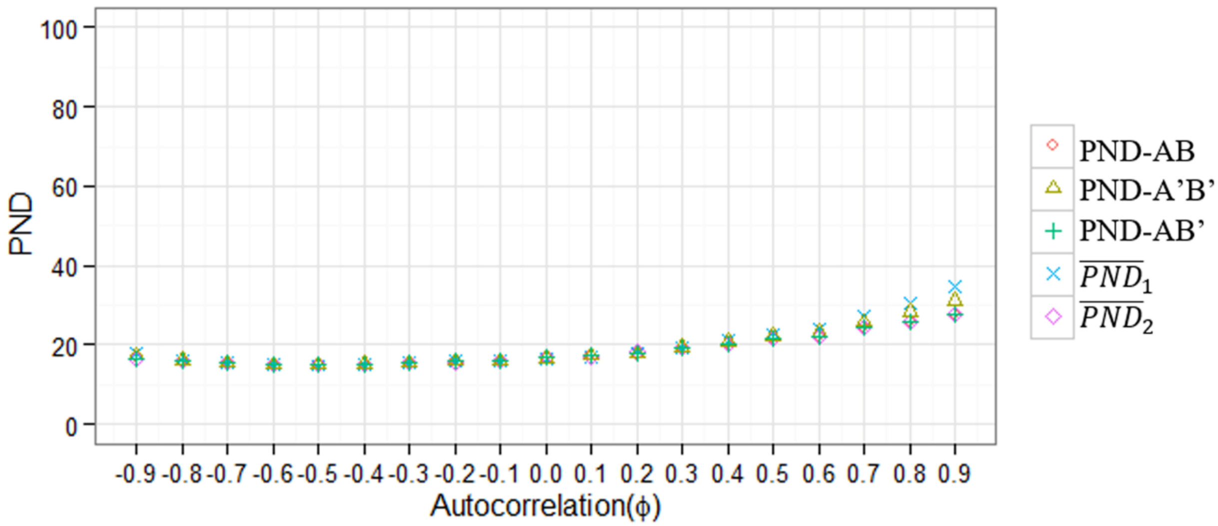


Figure 10. The effects of autocorrelation on PND ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

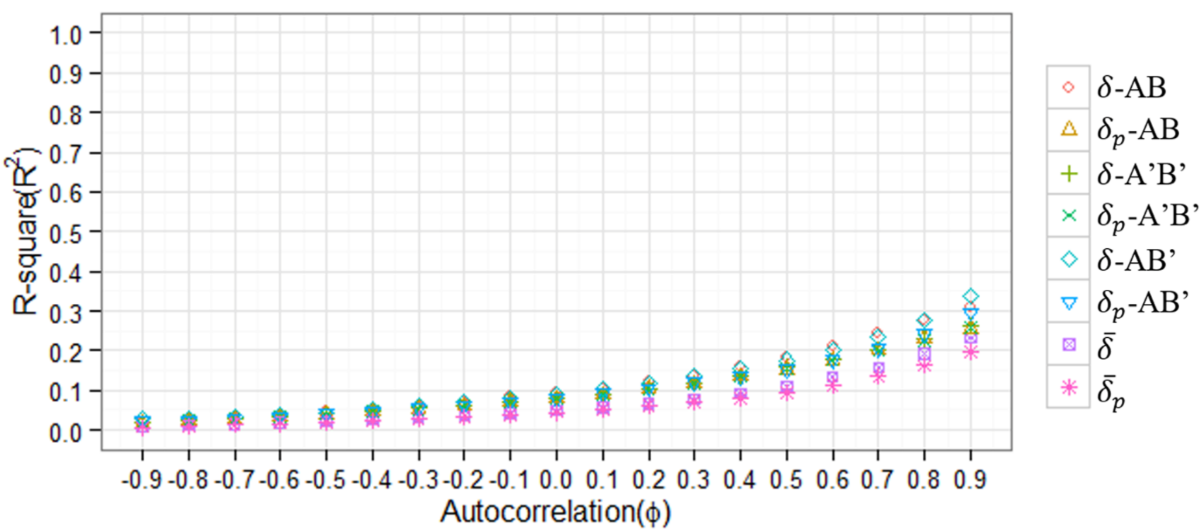


Figure 11. The effects of autocorrelation on SMDs converted to R^2 ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

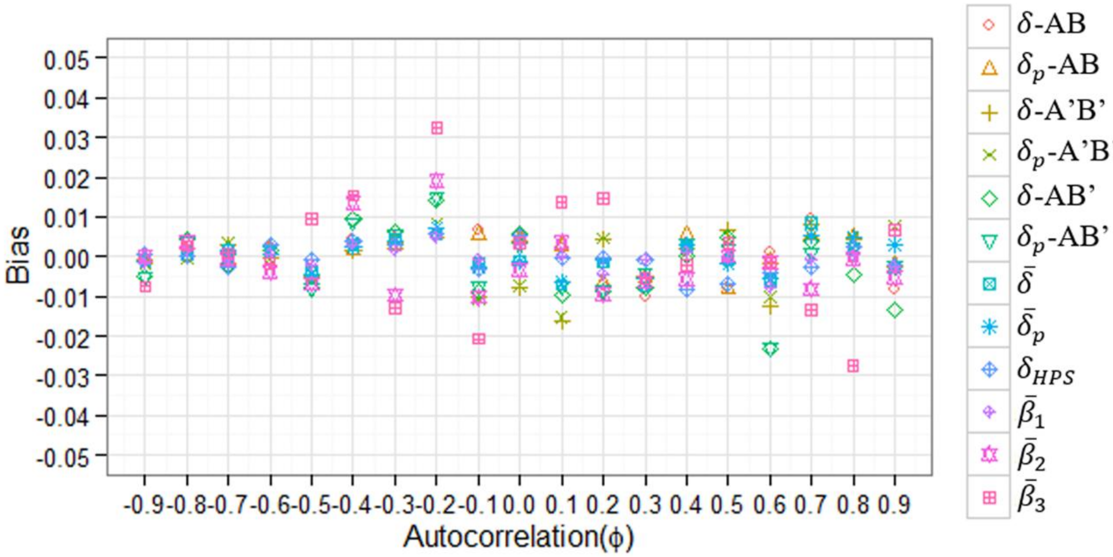


Figure 12. The effects of autocorrelation on the bias of partial regression coefficients and SMDs ($t_A = t_B = t_{A'} = t_{B'} = 5$) for no treatment effects.

Level Change Data Type

When only a change in level effect was present, partial regression coefficients, pooled SMD from Equation 5 for the first A and last B′ phases (δ_p-AB′), and Hedges et al.'s standardized mean difference (δ_{HPS}) ESs performed best in estimating the true effect (Figure 13, Figure 14). Nevertheless, both δ_p-AB′ and δ_{HPS} were affected by autocorrelation with the ESs becoming increasingly negatively biased at extreme degrees of autocorrelation. Unpooled SMD from Equation 4 for the first A and last B′ phases (δ-AB′) and standardized partial regression coefficients were positively biased in addition to being affected by autocorrelation. Other SMD ESs were biased estimates of the population effect. Similar to the previous condition, Allison and Gorman's R^2, sR^2s, and PND were consistently positively biased, with negative autocorrelation resulting in a smaller bias while positive autocorrelation resulted in larger bias (see Appendices E, F, and G).

Similarly, phase length also did not augment the relationship of ES statistics with autocorrelation for partial regression coefficients, SMDs, sR^2s, and PNDs (Appendix E, F, and Appendix G). Allison and Gorman's (1993) R^2 statistics were again affected by unequal phase lengths ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$) resulting in less variability among all four calculation methods (Appendix F). Likewise, the relationships between Allison and Gorman's R^2 statistics and autocorrelation were similar in equal phase length conditions.

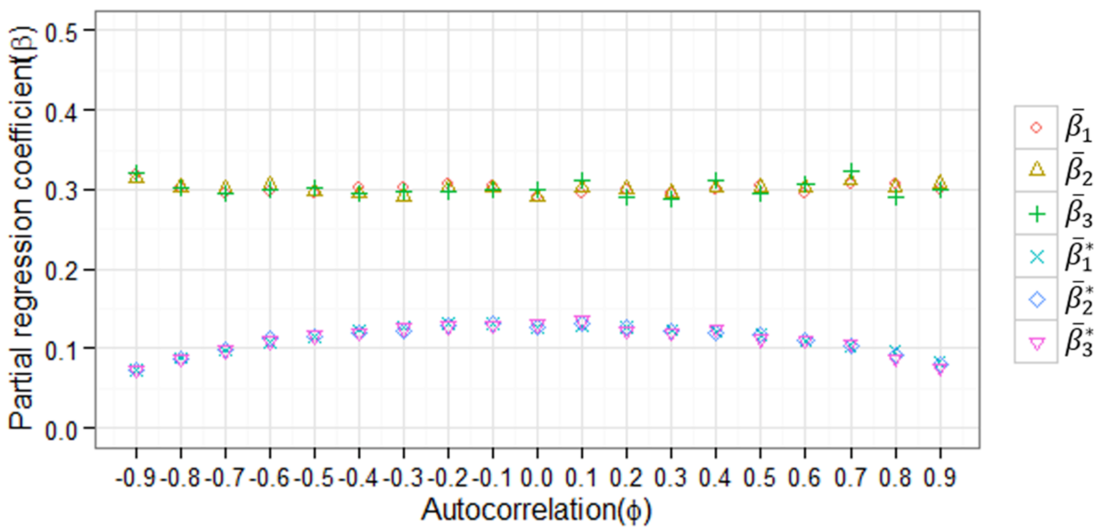


Figure 13. The effects of autocorrelation on standardized and unstandardized partial regression coefficients ($t_A = t_B = t_{A'} = t_{B'} = 5$). The true magnitude of effect to be estimated by partial regression coefficients is 0.30, and the standardized partial regression coefficients, 0.13.

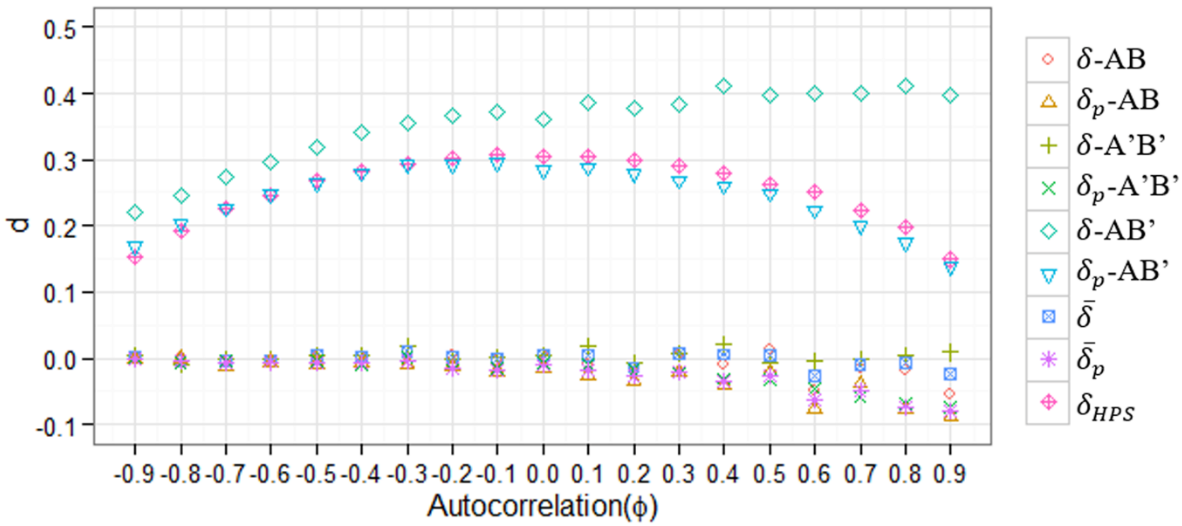


Figure 14. The effects of autocorrelation on standardized mean difference ($t_A = t_B = t_{A'} = t_{B'} = 5$). The true magnitude of effect to be estimated by the SMD ESs is 0.30.

Level Change and Extraneous Effects Data Type

When extraneous effects were present in addition to a change in level effect, an ideal ES would be sensitive and able to differentiate the increase in total effect. In the no autocorrelation condition, δ_{HPS} performed best in estimating the true effects in most data types and phase length combinations (Figure 15, Appendix Q) except for the $t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$ phase where the average of partial regression coefficients from Equation 1 ($\bar{\beta}_1$) performed best (Figure 16). Other partial regression coefficients and SMD ESs did not consistently and accurately estimate the appropriate effect changes in the presence of a time-based trend and a different slope in B and B′ phases. The partial regression coefficients tended to overestimate the true magnitude of effect when extraneous effects were present in the data. δ_p-AB′ also tended to overestimate the true magnitude of effect when extraneous effects were present in data patterns, more so in conditions where phase lengths were equal. As before, Allison and Gorman's R^2, sR^2s, and PND were consistently positively biased. The relationships of the ESs with autocorrelation when extraneous effects were present were comparable to those previously observed (see Appendices H to P). In conclusion, δ_{HPS} performed best in estimating the true effects in most conditions when no autocorrelation was present.

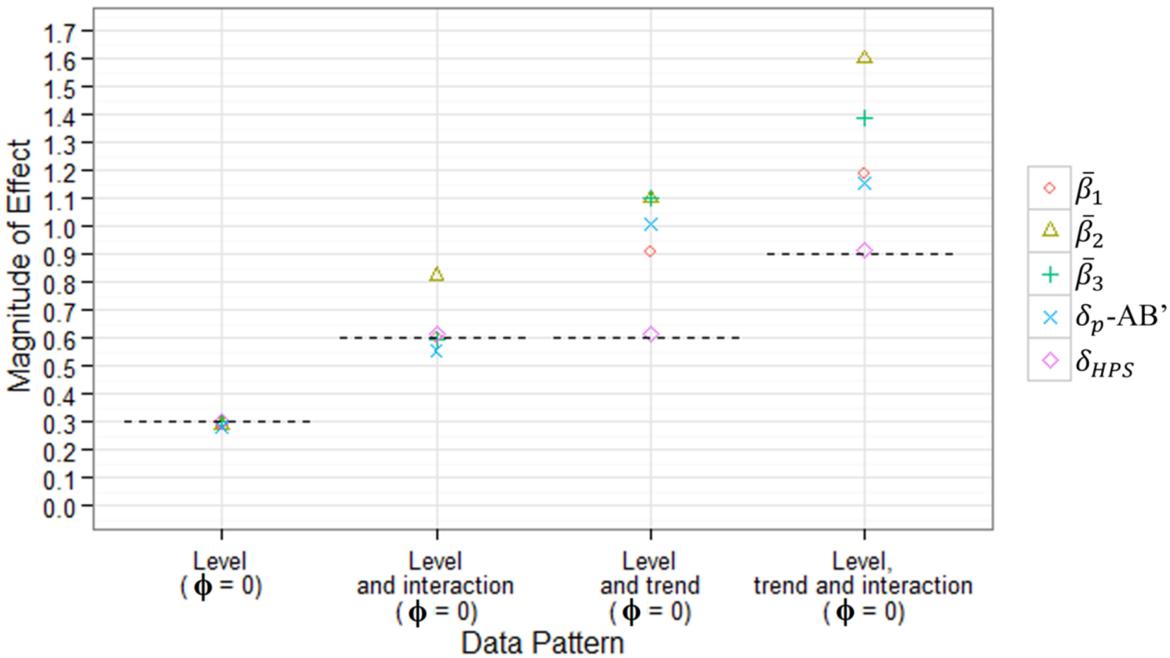


Figure 15. The effects of extraneous effects on partial regression coefficients and SMD ESs ($t_A = t_B = t_{A'} = t_{B'} = 5$). The true magnitude of effect to be estimated by a level change is 0.30, while level and interaction as well as level and trend change is 0.60, and level, trend, and interaction change is 0.90. There is no autocorrelation ($\varphi = 0$).

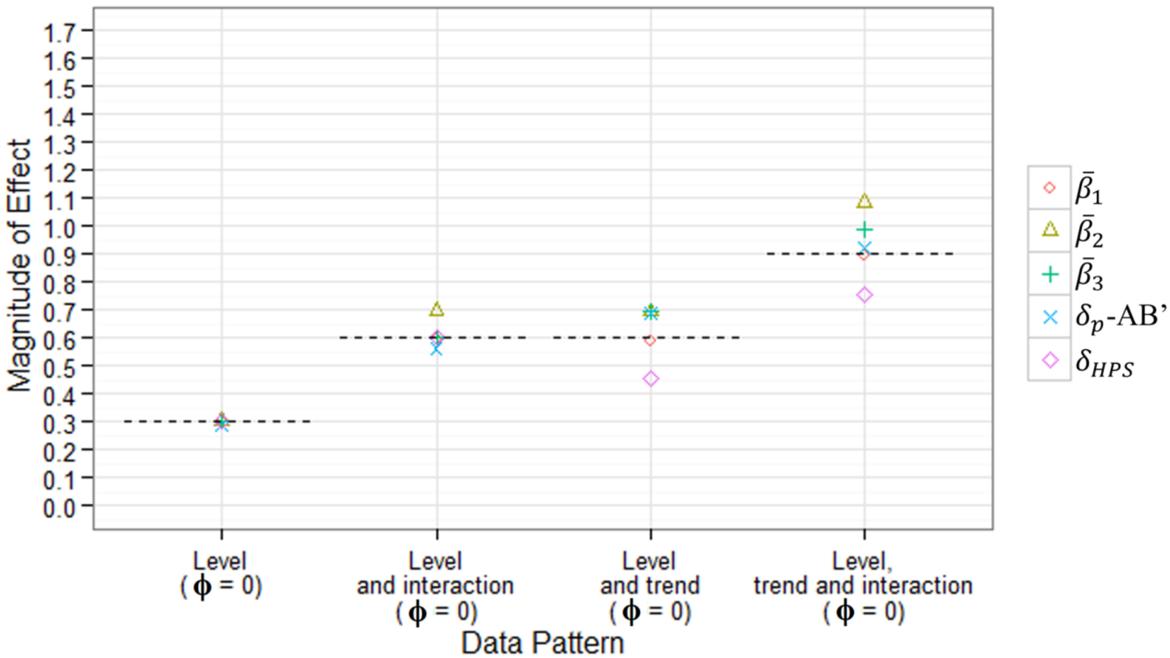


Figure 16. The effects of extraneous effects on partial regression coefficients and SMD ESs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$). The true magnitude of effect to be estimated by a level change is 0.30, while level and interaction as well as level and trend change is 0.60, and level, trend, and interaction change is 0.90. There is no autocorrelation ($\varphi = 0$).

When small to medium degrees of autocorrelation were present (e.g. φ = 0.3 and φ = -0.3), δ_{HPS} still performed best in estimating the true effect in even phases while $\bar{\beta}_1$ performed best when phases are uneven (see Appendices R to T). However, when medium to large degrees of autocorrelation were present (e.g. φ = 0.6 and φ = -0.6), δ_{HPS} consistently underestimated the true effect in all conditions (Figure 17, Figure 18, Appendices U to W). $\bar{\beta}_1$, however, was consistent in estimating the true effect for uneven phases conditions, and for even phases conditions when only a level change in addition to a level change and interaction effects type were present (Figure 17, Figure 18, Appendices U to W).

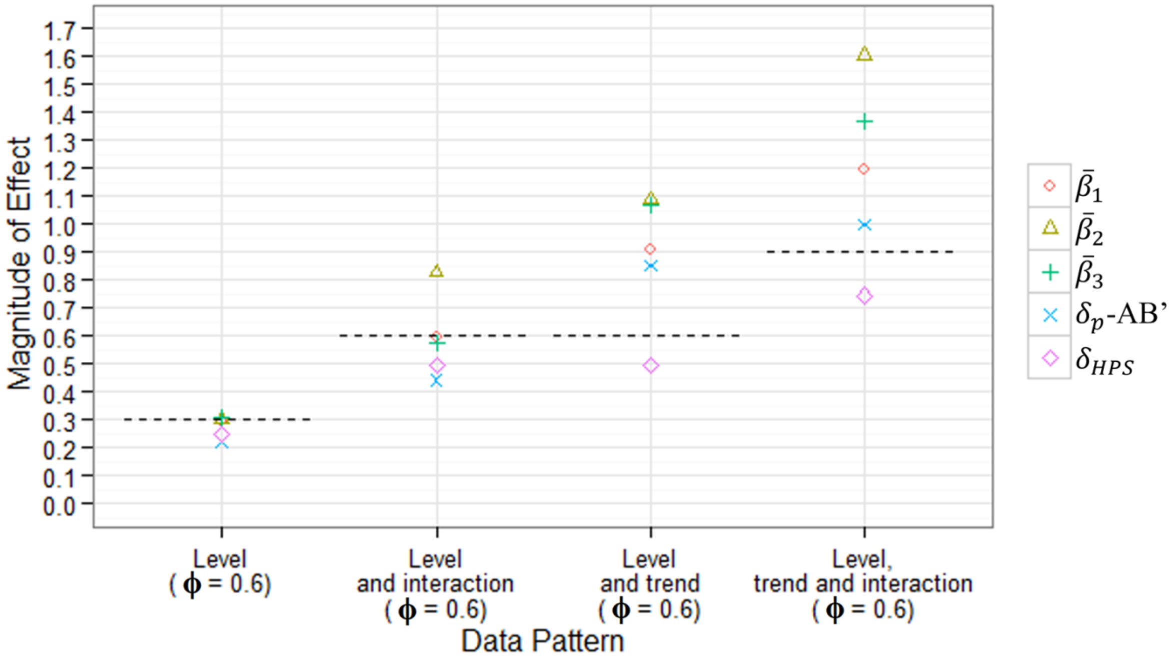


Figure 17. The effects of extraneous effects on partial regression coefficients and SMD ESs ($t_A = t_B = t_{A'} = t_{B'} = 5$). The true magnitude of effect to be estimated by a level change is 0.30, while level and interaction as well as level and trend change is 0.60, and level, trend, and interaction change is 0.90. Autocorrelation degree is $\varphi = 0.6$.

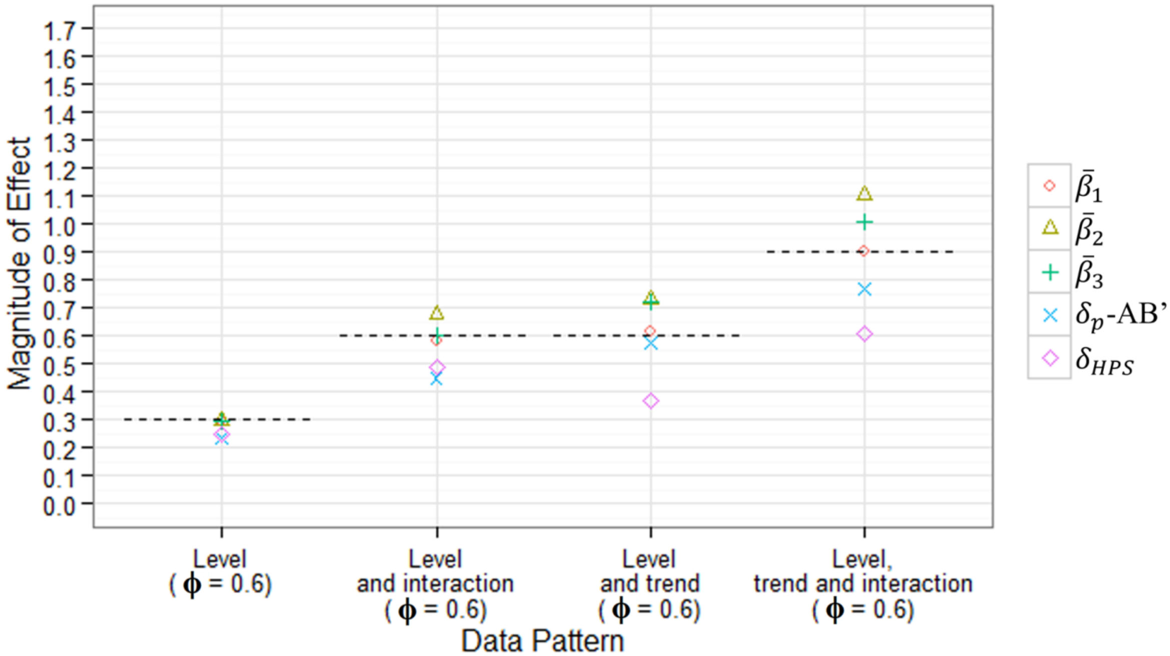


Figure 18. The effects of extraneous effects on partial regression coefficients and SMD ESs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$). The true magnitude of effect to be estimated by a level change is 0.30, while level and interaction as well as level and trend change is 0.60, and level, trend, and interaction change is 0.90. Autocorrelation degree is $\varphi = 0.6$.

Discussion

The present thesis evaluated the statistical properties of various ES statistics for a four-phase ABA′B′ SCED, using simulated data which exhibited autocorrelation, a time-based trend, and different slopes for time-trends within the phases. SCEDs are a set of experimental designs that produce data similar to that of a short and interrupted time-series where one of the main challenges is that the observations may not be independent due to autocorrelation. Consequently, to date there is a lack of consensus regarding the optimal ESs for application to meta-analyses. Since ESs are the primary data for a meta-analysis, this is a preliminary challenge and requirement for further advancement for meta-analyses of SCEDs. The best performing ES should be sensitive enough to accurately estimate an effect in short phase lengths and not affected by autocorrelation. Among all the ESs evaluated, Hedges, Pustejovsky, and Shadish's (2012) standardized mean difference (δ_{HPS}) performed best according to these criteria. Nevertheless, δ_{HPS} was affected by medium to extreme degrees of autocorrelation and slightly affected by unequal phase lengths. The results of the present thesis are discussed in comparison with the findings of Manolov and Solanas (2008), followed by the effect of autocorrelation on standardized ESs, the utility of δ_{HPS} in an amalgamation of both SCEDs and group designs in a meta-analysis, and directions for future research.

Manolov and Solanas (2008) investigated the performance of PND, unpooled SMD from Equation 4 (δ-AB) in R^2 metric, pooled SMD from Equation 5 (δ_p-AB) in R^2 metric, and Allison and Gorman's R^2 (AG-R^2-AB), for a two-phase AB SCED. In this thesis, similar results were obtained for the ESs Manolov and Solanas (2008) investigated when the two-phase AB SCED was extended to a four-phase ABA′B′ SCED. Among all the ESs investigated by Manolov and Solanas (2008), the PND seemed to have performed best because by definition the R^2 ESs are bounded at zero and thus are consistently positively biased. In addition to PND and ESs in the R^2 metric investigated by Manolov and Solanas (2008), the present thesis also investigated partial regression coefficients and their standardized counterparts, as well as SMD ESs and δ_{HPS}. PND was less sensitive to the different types of effects that may be present when compared to partial regression coefficients, SMD and δ_{HPS} ESs. The present thesis found δ_{HPS} to be the best performing ES for ABA′B′ SCEDs for SCED data with small to moderate degrees of autocorrelation.

In addition to δ_{HPS}, partial regression coefficients also performed well across a variety of conditions. Specifically, the unstandardized partial regression coefficients performed better than δ_{HPS} in the presence of moderate to high autocorrelation. Since all regression-based ESs was estimated via regression models that implicitly assumed independence of errors, the ESs were a product of regression models that were misspecified. Nevertheless, autocorrelation did not affect unstandardized estimates of the partial regression coefficients. Unstandardized partial regression coefficients were not rescaled, thus did not incorporate information from the estimated variance. Perhaps autocorrelation biased the variance estimates and hence produced biased standardized partial regression coefficients which had an association with autocorrelation shaped like an inverted ‘U’. Furthermore, most of the other standardized ESs evaluated in this thesis exhibited this similar relationship with autocorrelation. This result is encouraging for researchers looking to combine unstandardized partial regression coefficients and researchers that prefer to use and report unstandardized ESs. Additionally, when the scales of measurement are similar and meaningful, unstandardized ESs are advantageous for interpretation.

In certain conditions, some SMD ESs failed to detect an effect when there was indeed a non-zero true effect. This result occurs because SCEDs generally produce data with a short series coupled with potentially large variability among phases. The SMD ESs that failed to find non-zero effects in the present thesis may be able to obtain more accurate estimates with SCED data of longer series and smaller variability among phases.

One of the goals of the thesis was to find an ES for SCEDs which is comparable to group-based experimental design ESs. Two ESs are comparable when they estimate equivalent parameters. As previously described, SCEDs are a set of experimental designs that may include replications across units, settings, and practitioners. Thus a SCED study may include one or more units that are subjected to the same SCED design. Among all the ESs evaluated, δ_{HPS} is the only ES designed to be comparable to Cohen's d (see Hedges et al., 2012). Given its favorable performance in the present thesis, the utility of δ_{HPS} in the amalgamation of both SCEDs and group designs in a meta-analysis is no doubt an important advantage. δ_{HPS} is designed for use with SCED studies consisting of more than two units. If the SCED study consists of more than two units, δ_{HPS} is recommended as the ES of choice.

Future studies may explore δ_{HPS} in terms of its statistical properties with different data patterns (e.g. data with only a time-based trend effect), different autocorrelation processes, in addition to δ_{HPS}'s performance when combined with ESs obtained from group designs in a meta-analysis. Nevertheless, as evidenced by the present simulation SCED researchers should report δ_{HPS} whenever possible as δ_{HPS} was one of the best performing ESs. Reporting partial regression coefficients is also beneficial, especially with designs utilizing unequal phase lengths.

References

Allen, K. D., Friman, P. C., & Sanger, W. G. (1992). Small n research designs in reproductive toxicology. *Reproductive Toxicology, 6(2),* 115-21.

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy, 31(6),* 621-631.

Barker, J. B., Mellalieu, S. D., McCarthy, P. J., Jones, M. V., & Moran, A. (2013). Special issue on single-case research in sport psychology. *Journal of Applied Sport Psychology, 25*(1), 1-3. doi:http://dx.doi.org/10.1080/10413200.2012.729378

Barlow, D. H., Nock, M. K., & Hersen, M. (2008). *Single case experimental designs: Strategies for studying behavior for change* (3rd ed.). Boston: Pearson.

Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention, 2(3),* 129-141. doi:10.1080/17489530802446302

Berliner, D. C. (2002). Comment: Educational research--the hardest science of all. *Educational Researcher, 31*(8), 18-20.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Research Design and Analysis: New directions for psychology and education.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Burns, M. K. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education, 21(3),* 175-184. doi:http://dx.doi.org/10.1007/s10864-012-9158-9

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33*(4), 269-285.

Center, B. A., Skiba, R. J., & Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education, 19*(4), 387-400.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*(1), 12-37. doi:http://dx.doi.org/10.1177/0163278702025001003

Evidence-based practice in psychology. (2006). *American Psychologist, 61(4)*, 271-285. doi:http://dx.doi.org/10.1037/0003-066X.61.4.271

Ferron, J. (2002). Reconsidering the use of the general linear model with single-case data. *Behavior Research Methods, Instruments & Computers, 34*(3), 324-331. doi:http://dx.doi.org/10.3758/BF03195459

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York: Routledge.

Gingerich, W. J. (1984). Meta-analysis of applied time-series data. *Journal of Applied Behavioral Science, 20*(1), 71-79.

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment, 5*(2), 141-154.

Guyatt, G. H., Haynes, R. B., Jaeschke, R. Z., Cook, D. J., et al. (2000). Users' guides to the medical literature: XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. *Journal of the American Medical Association, 284*(10), 1290-6.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*(3), 224-239. doi:http://dx.doi.org/10.1002/jrsm.1052

Horner, R. H., Carr, E. G., Halle, J., McGee, G., & al, e. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165-179.

Johnston, M. V., & Smith, R. O. (2010). Single subject designs: Current methodologies and future directions. *Occupational Therapy Journal of Research: Occupation, Participation and Health*, *30*(1), 4-10. doi:http://dx.doi.org/10.3928/15394492-20091214-02

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

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137-152. doi:http://dx.doi.org/10.1037/a0028086

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=229.

Lane, K. L., & Carter, E. W. (2013). Reflections on the special issue: Issues and advances in the meta-analysis of single-case research. *Remedial and Special Education*, *34*(1), 59-61. doi:http://dx.doi.org/10.1177/0741932512454582

Ma, H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification*, *30*(5), 598-617. doi:http://dx.doi.org/10.1177/0145445504272974

Macgowan, M. J., & Wong, S. E. (2014). Single-case designs in group work: Past applications, future directions. *Group Dynamics: Theory, Research, and Practice, 18*(2), 138-158. doi:http://dx.doi.org/10.1037/gdn0000003

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985–2009. *Exceptionality, 19*(2), 109-135. doi:http://dx.doi.org/10.1080/09362835.2011.565725

Manolov, R., & Solanas, A. (2008). Comparing N = 1 effect size indices in presence of autocorrelation. *Behavior Modification,32*(6), 860-875. doi:http://dx.doi.org/10.1177/0145445508318866

Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation, 19*(6), 904-927. doi:http://dx.doi.org/10.1080/09602010903040691

Robb, S. L. (2013). Editorial: The power of the pilot. *Journal of Music Therapy, 50*(1), 3-5.

Robey, R. R., Schultz, M. C., Crawford, A. B., & Sinner, C. A. (1999). Single-subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology, 13*(6), 445-473.

Ross, S. G., & Begeny, J. C. (2014). Single-case effect size calculation: Comparing regression and non-parametric approaches across previously published reading intervention data sets. *Journal of School Psychology, 52*(4), 419-431. doi:http://dx.doi.org/10.1016/j.jsp.2014.06.003

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*(1), 9-19.

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

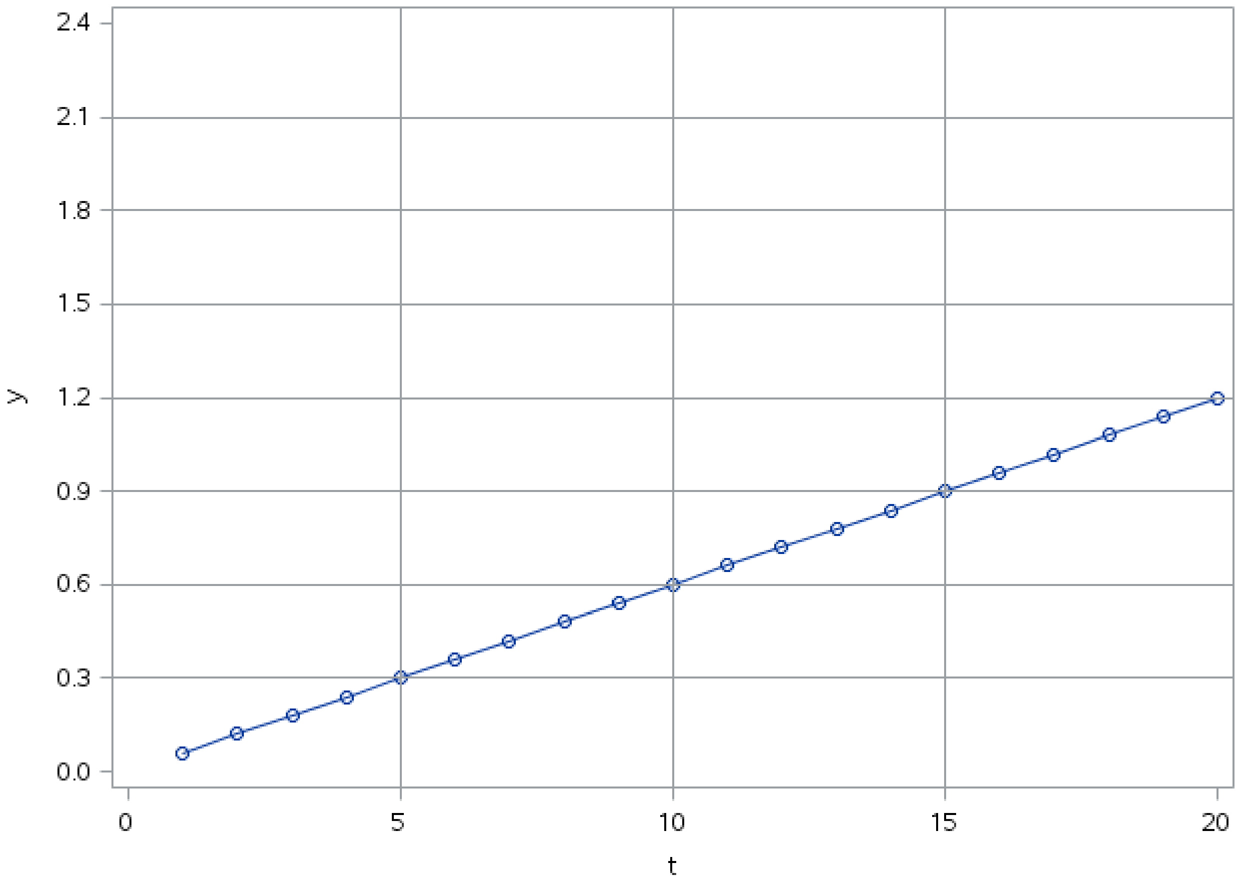
Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, (113), 95-109.

Smith G.C. & Pell J.P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal*, 327(7429), 1459-61.

Appendices

Appendix A

The effect of a time-based trend.



The population effect for AB phases is:

$$\frac{0.36 + 0.42 + 0.48 + 0.54 + 0.60}{5} - \frac{0.06 + 0.12 + 0.18 + 0.24 + 0.30}{5}$$

= 0.48 – 0.18 = 0.30

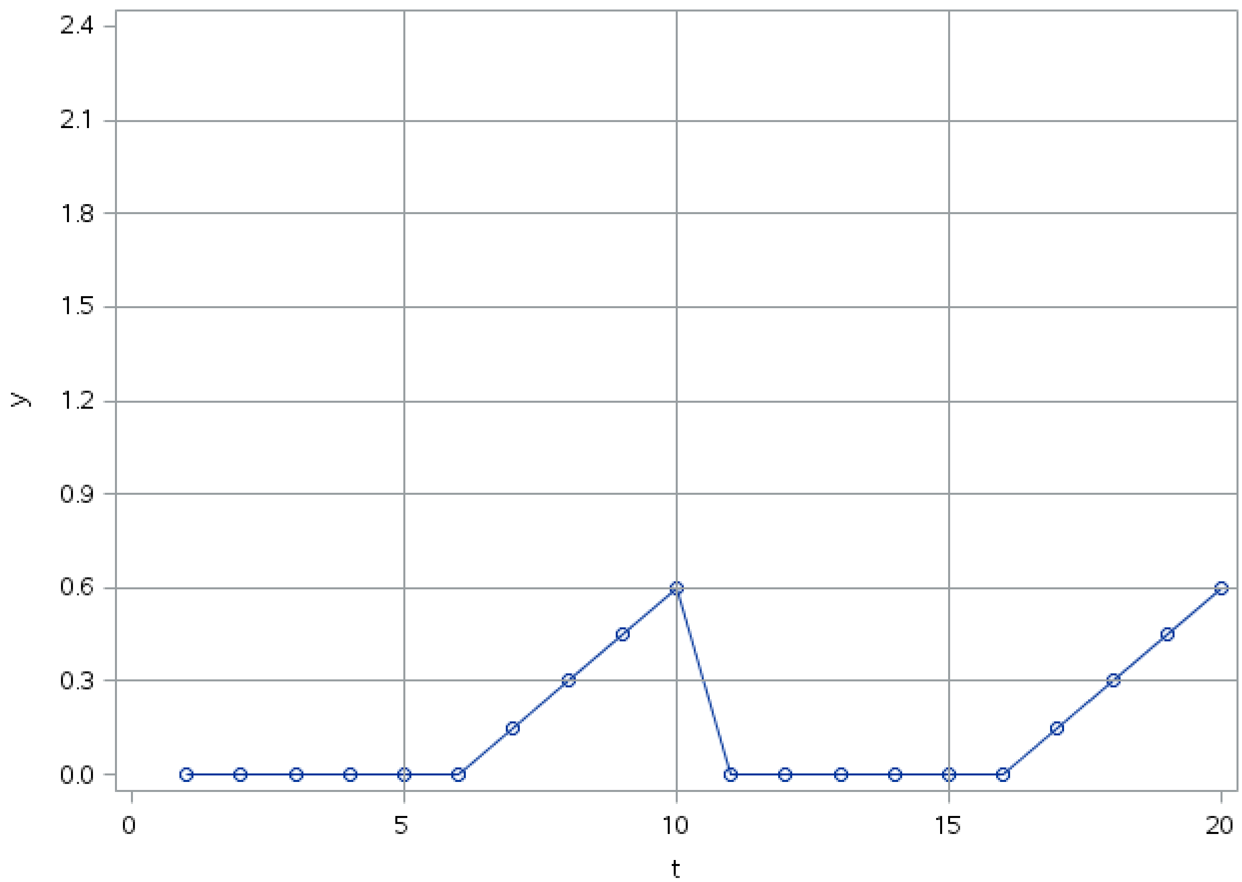
The population effect for A′B′ phases is:

$$\frac{0.98 + 1.02 + 1.08 + 1.14 + 1.20}{5} - \frac{0.66 + 0.72 + 0.78 + 0.84 + 0.90}{5}$$

= 1.08 – 0.78 = 0.30

Appendix B

The effect of a slope within phases B or B′.



The population effect for AB phases is:

$$\frac{0 + 0.15 + 0.30 + 0.45 + 0.60}{5} - \frac{0}{5}$$

= 0.30 – 0 = 0.30

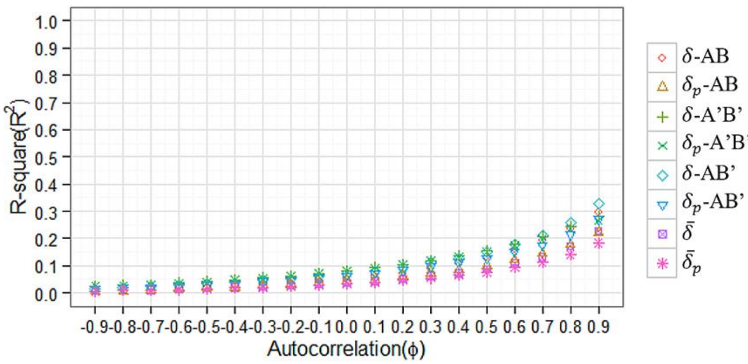
The population effect for A′B′ phases is:

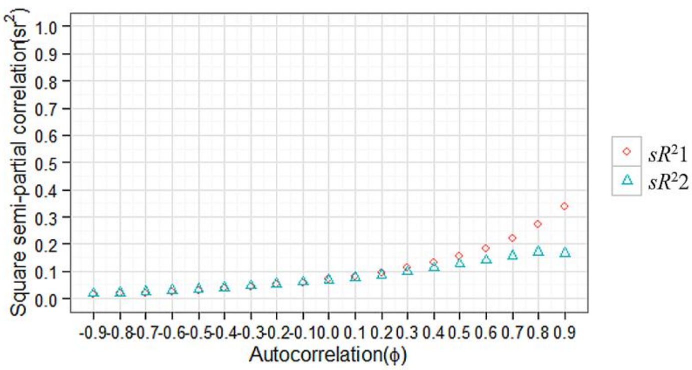
$$\frac{0 + 0.15 + 0.30 + 0.45 + 0.60}{5} - \frac{0}{5}$$

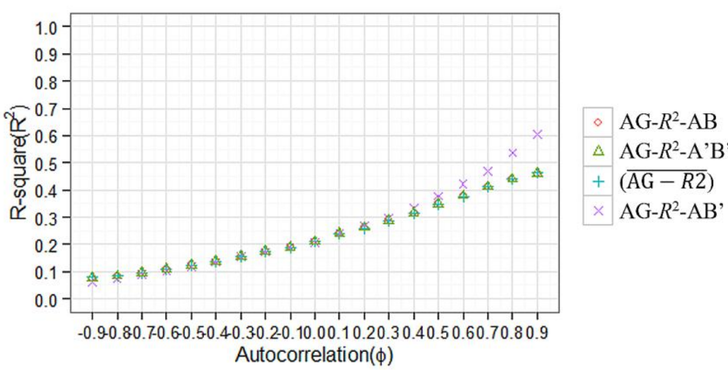
= 0.30 – 0 = 0.30

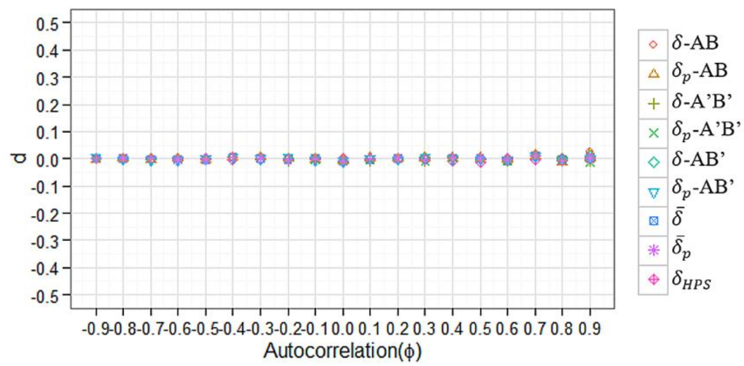
Appendix C

No effects data type, autocorrelation graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)

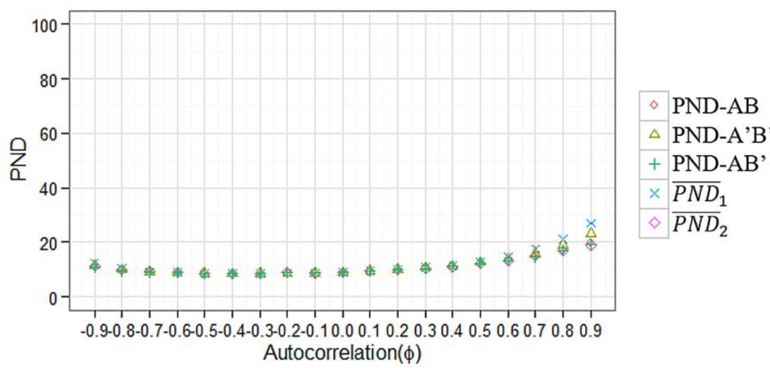




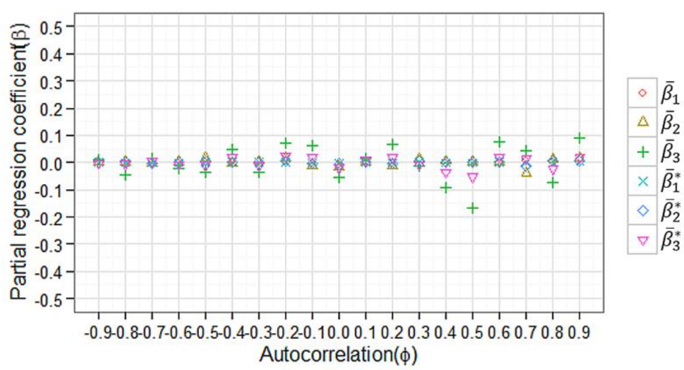




d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}



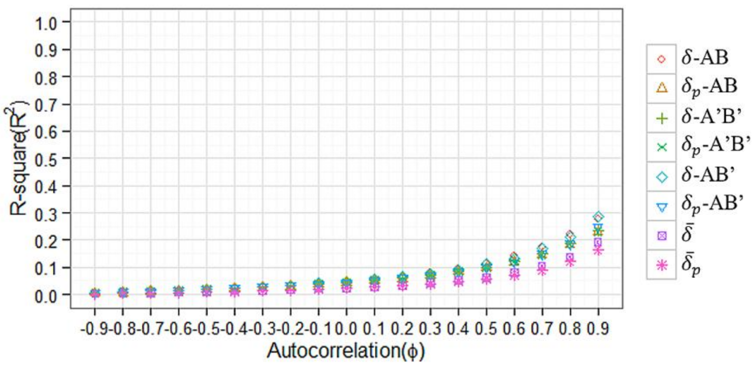
PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

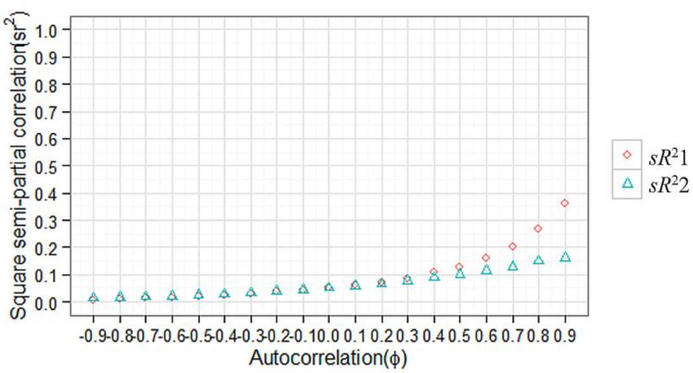


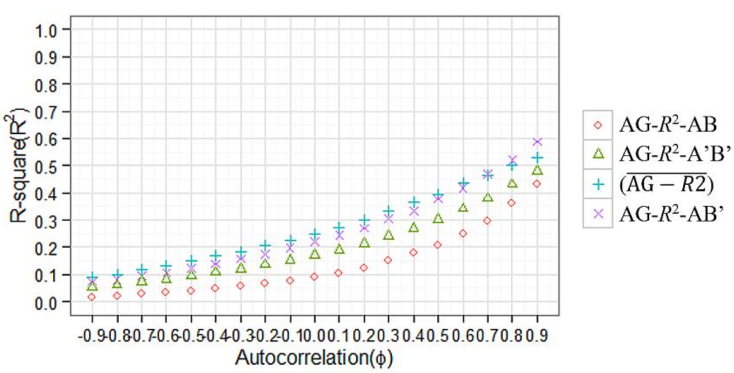
Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

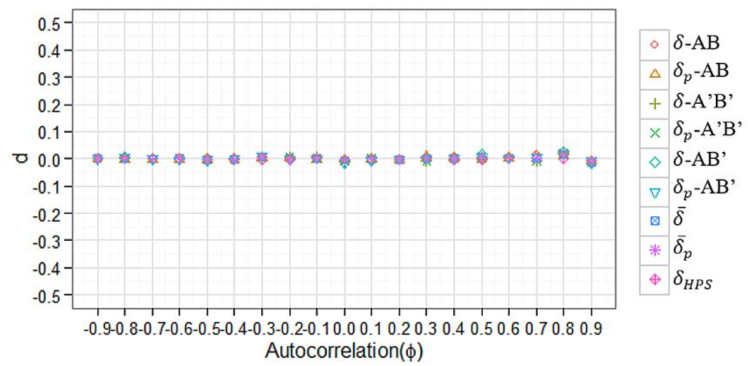
Appendix D

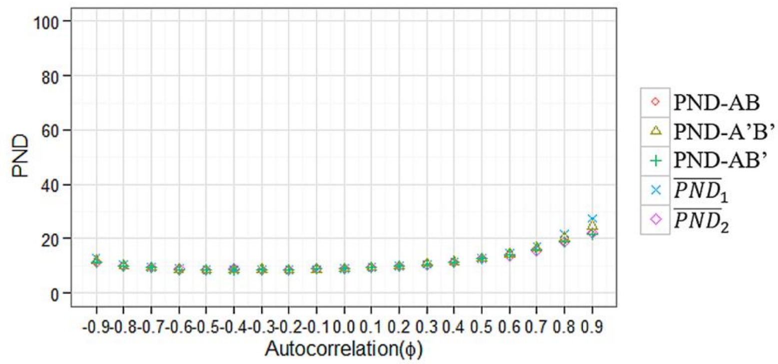
No effects data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

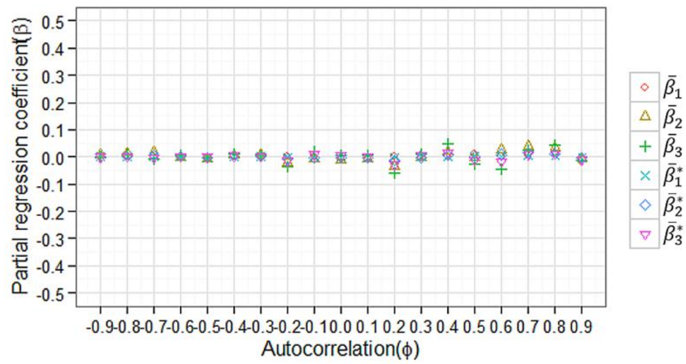






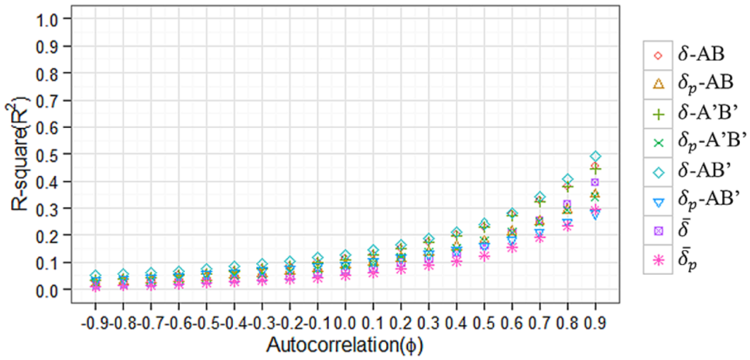

d
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

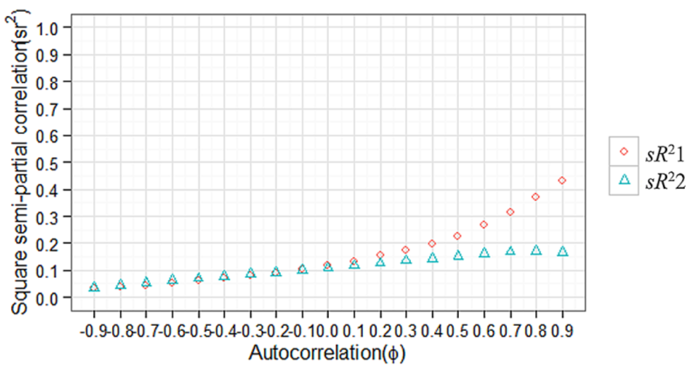

PND
100
80
60
40
20
0
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

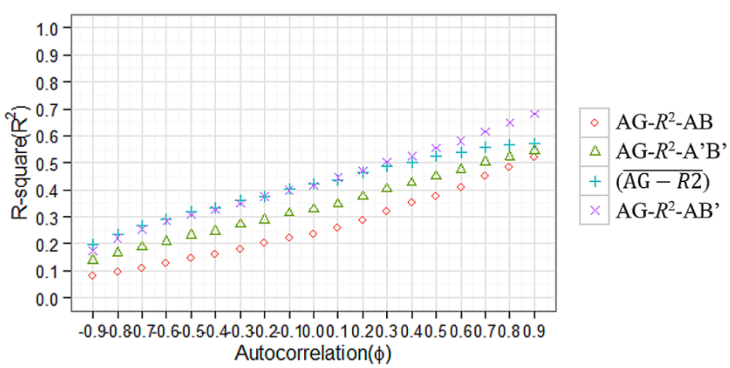

Partial regression coefficient(β)
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

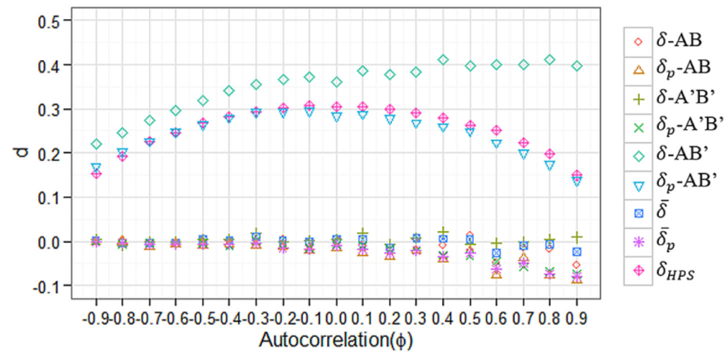
Appendix E

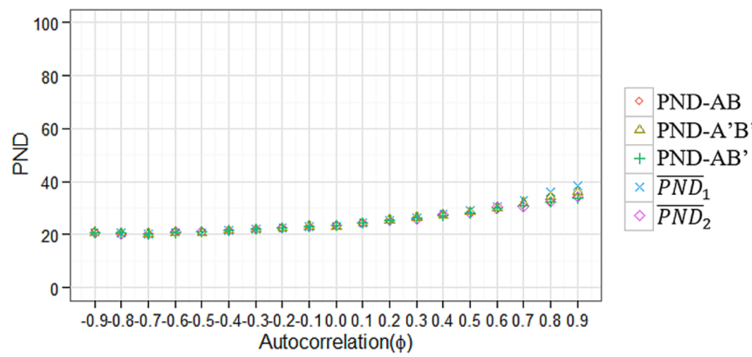
Level change data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)

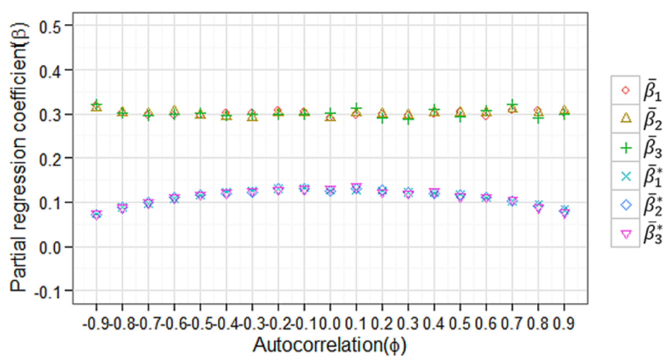






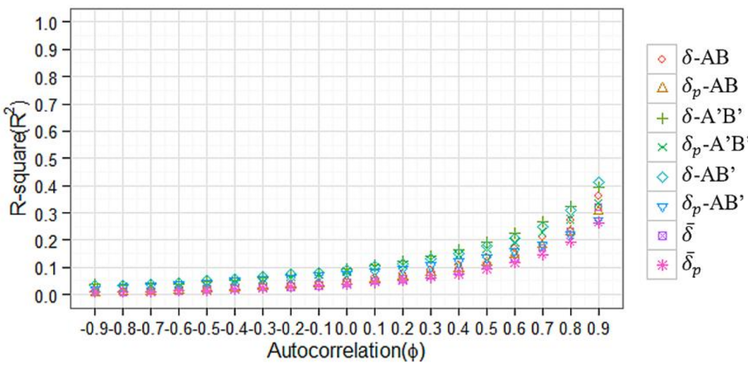

d
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

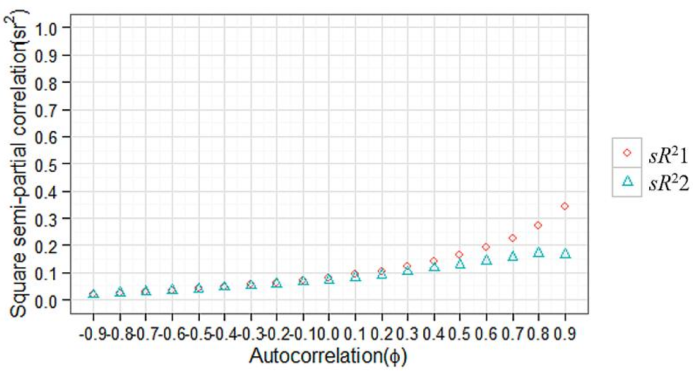

PND
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

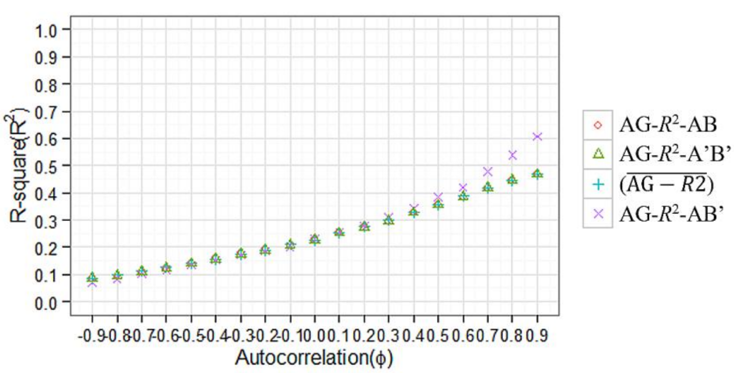

Partial regression coefficient(β)
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

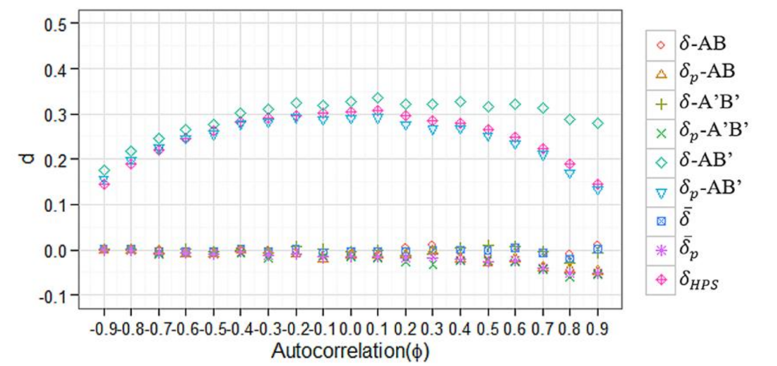
Appendix F

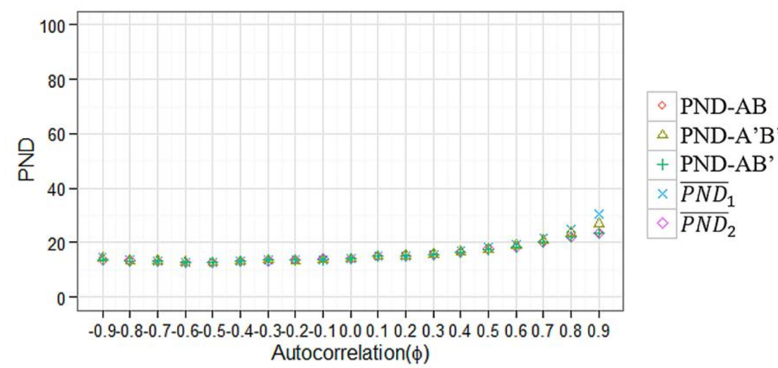
Level change data type, autocorrelation graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)

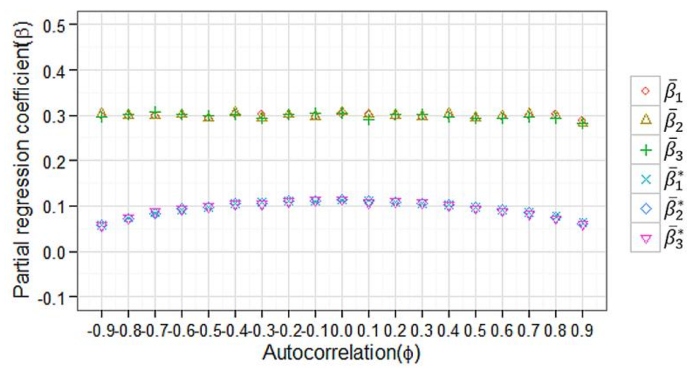






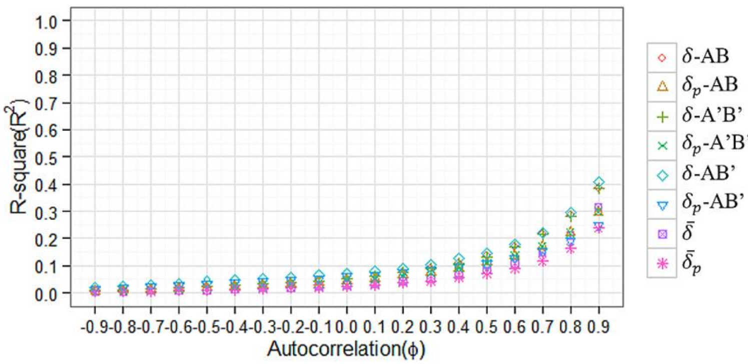

d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

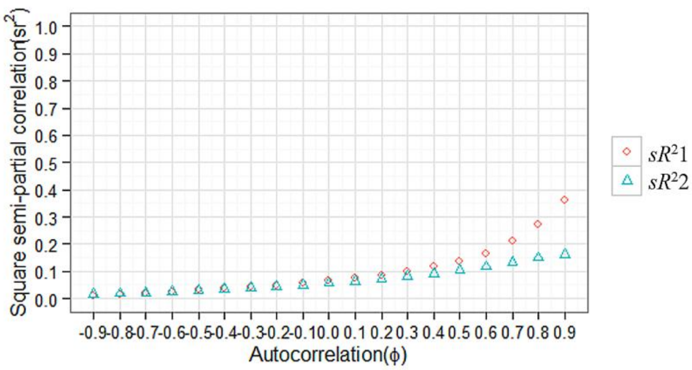

PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

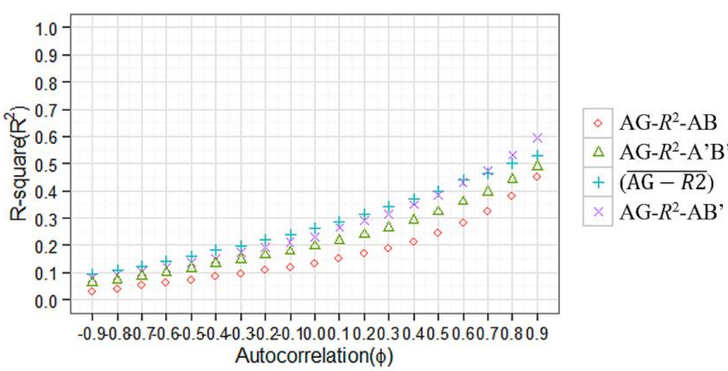

Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

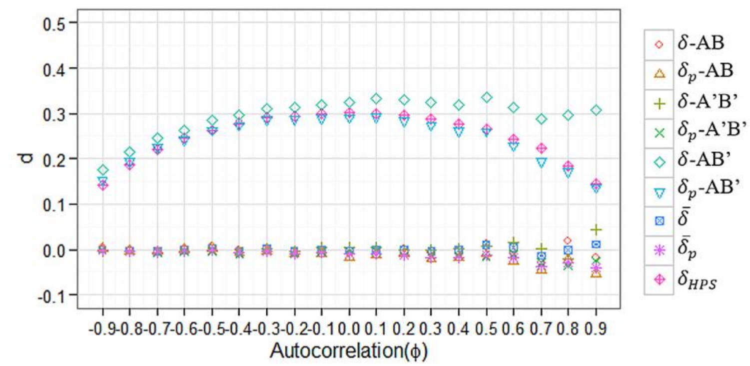
Appendix G

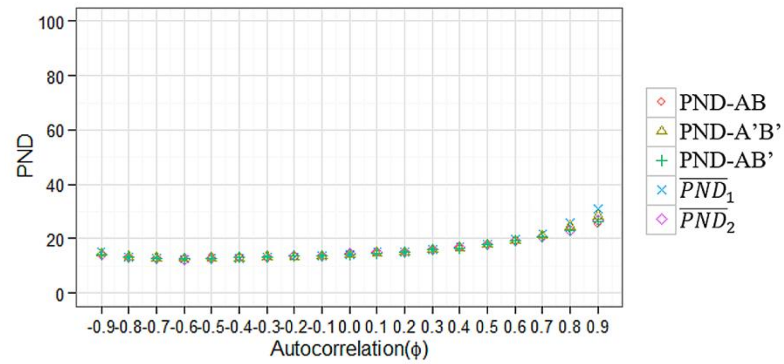
Level change data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

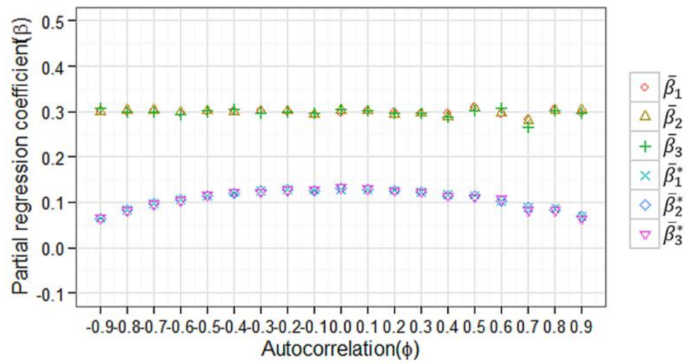






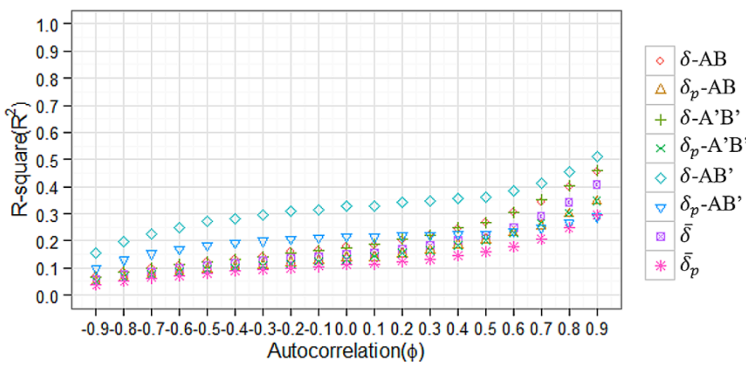

d
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

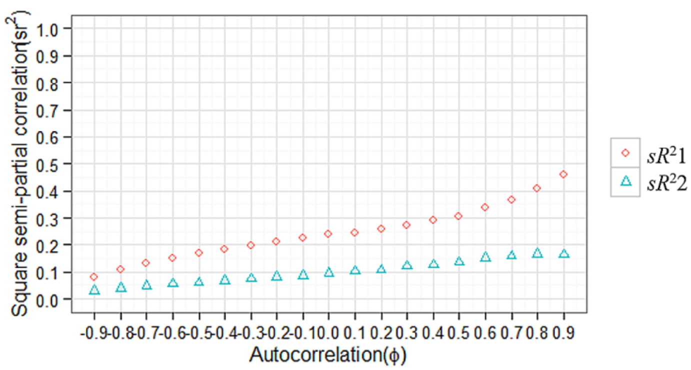

PND
100
80
60
40
20
0
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

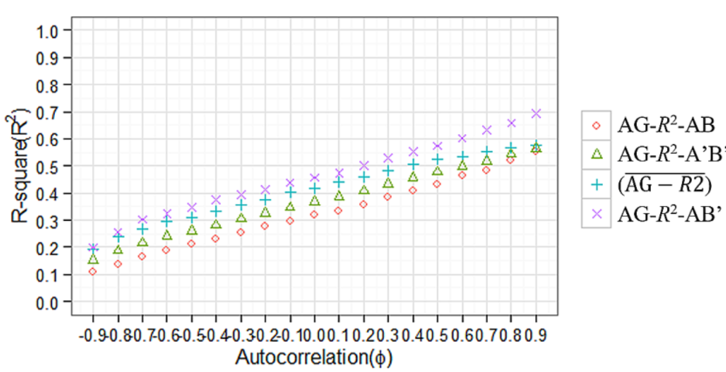

Partial regression coefficient(β)
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

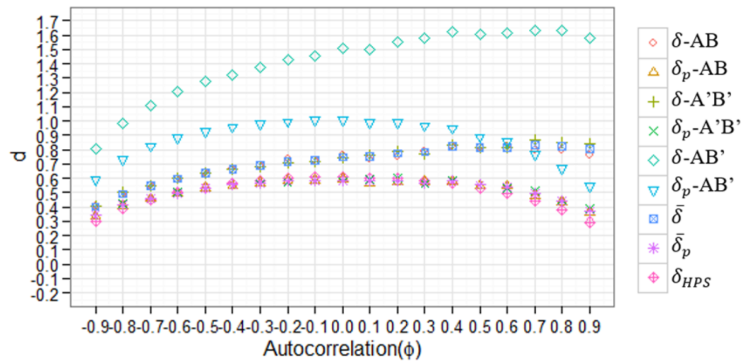
Appendix H

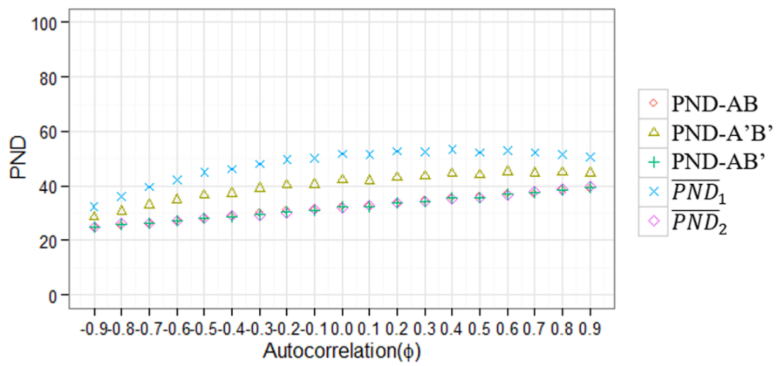
Level and trend data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)

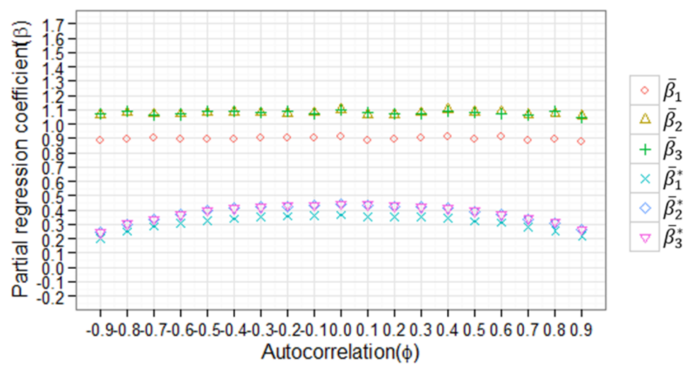






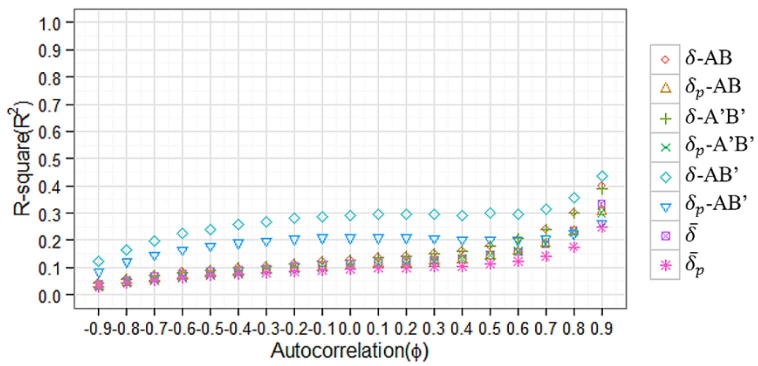

d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

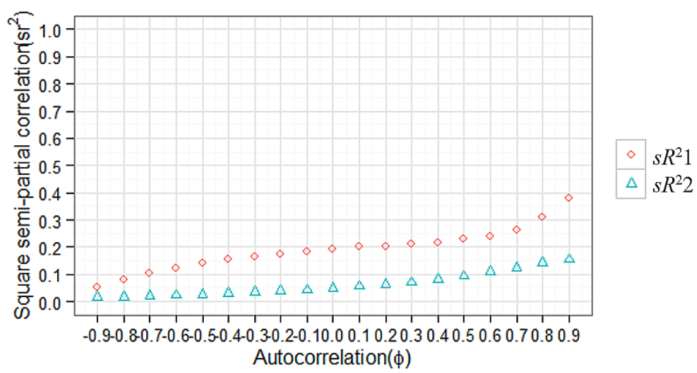

PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

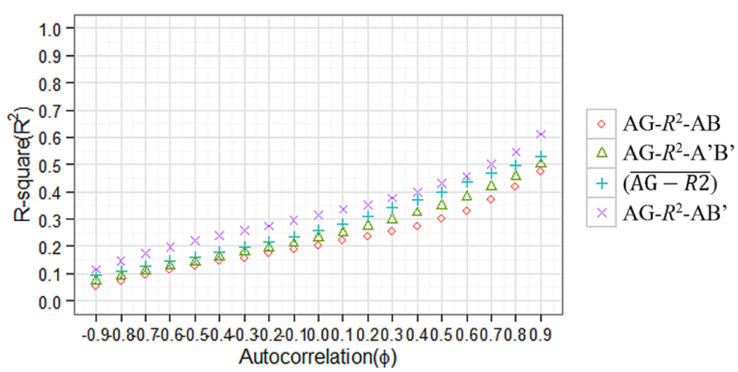

Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

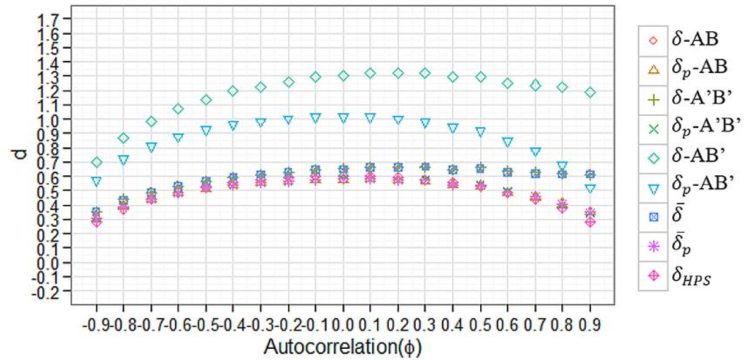
Appendix I

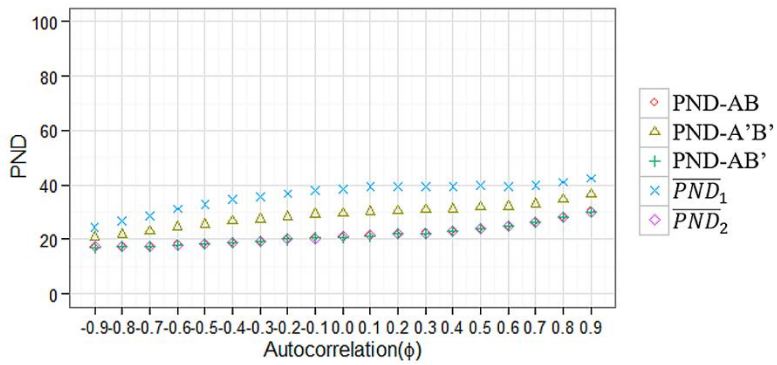
Level and trend data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

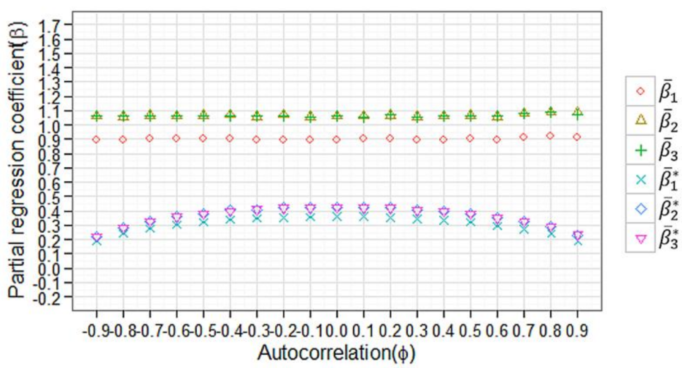






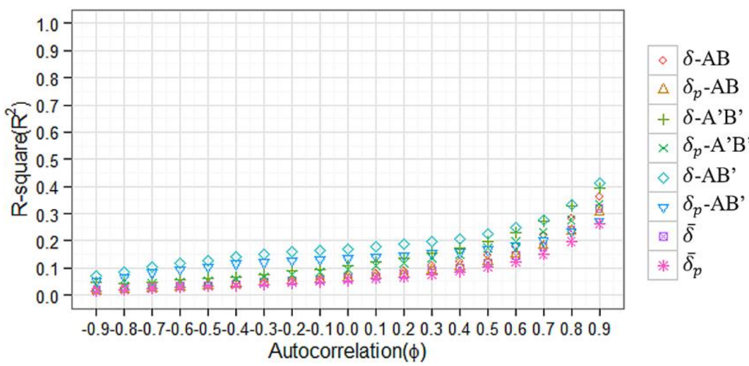

d
Autocorrelation(ϕ)
δ-AB
δp-AB
δ-A'B'
δp-A'B'
δ-AB'
δp-AB'
δ̄
δ̄p
δHPS

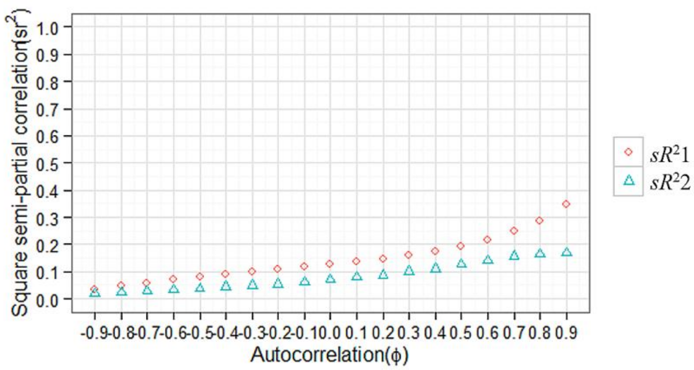

PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
PND1
PND2

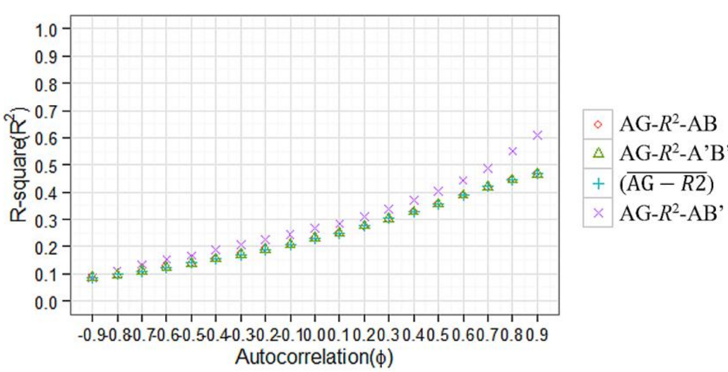

Partial regression coefficient(β)
Autocorrelation(ϕ)
β̄1
β̄2
β̄3
β̄1*
β̄2*
β̄3*

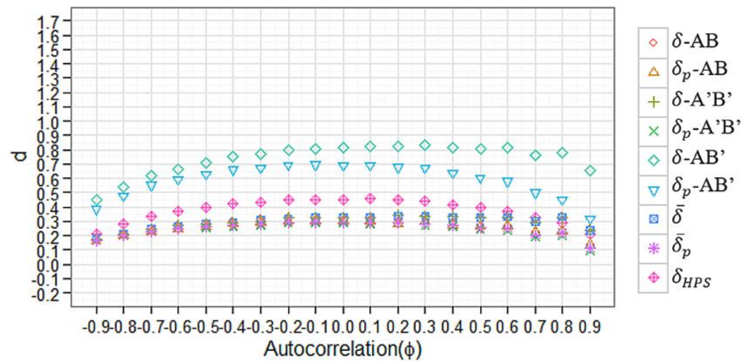
Appendix J

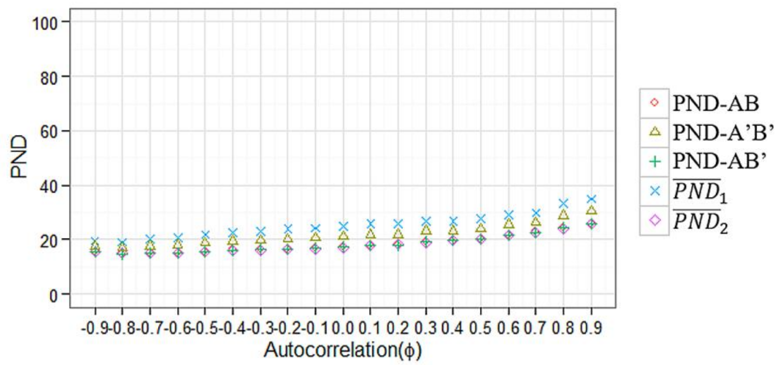
Level and trend data type, autocorrelation graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)

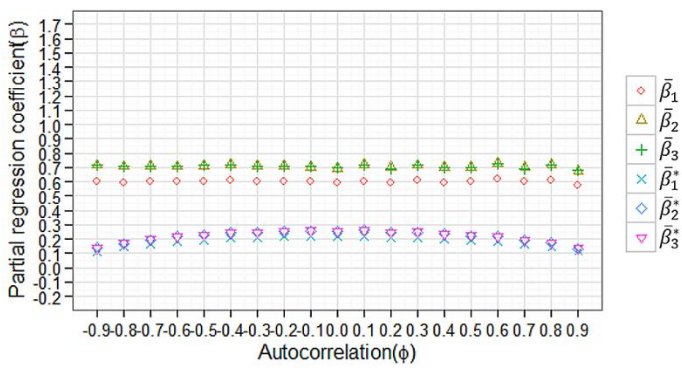






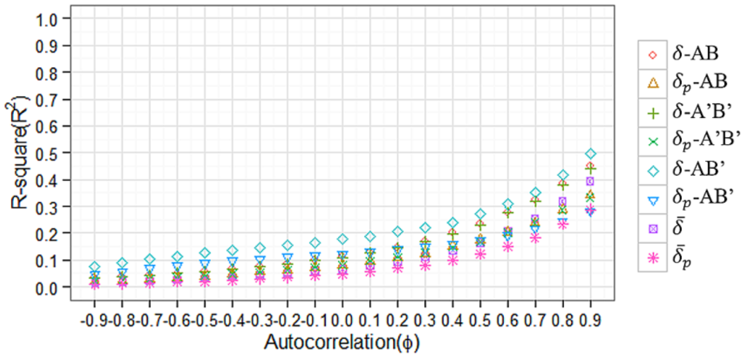

d
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

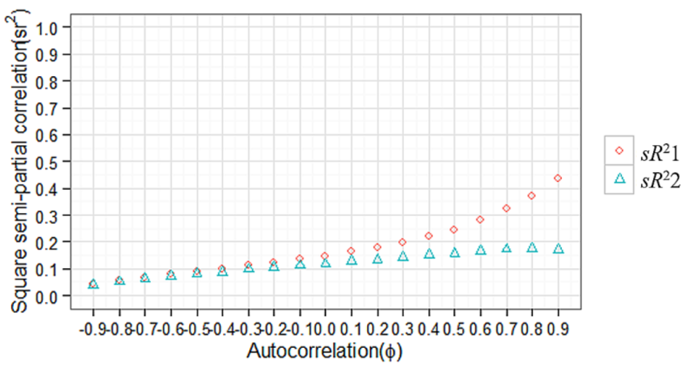

PND
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

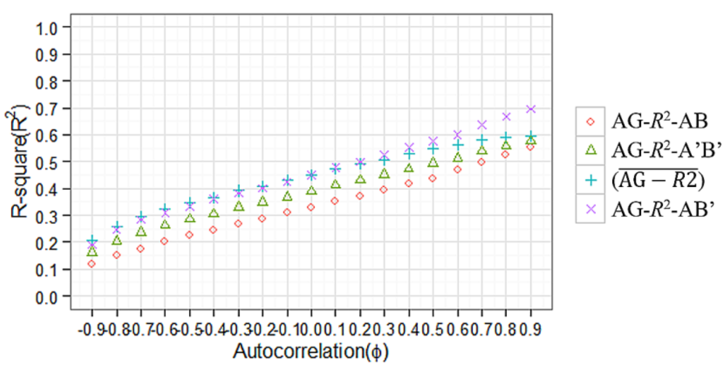

Partial regression coefficient(β)
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

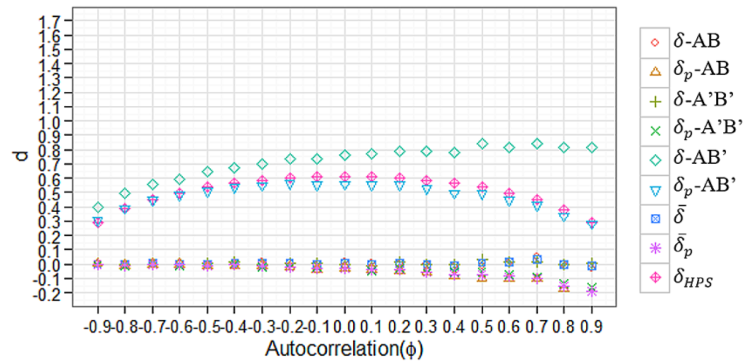
Appendix K

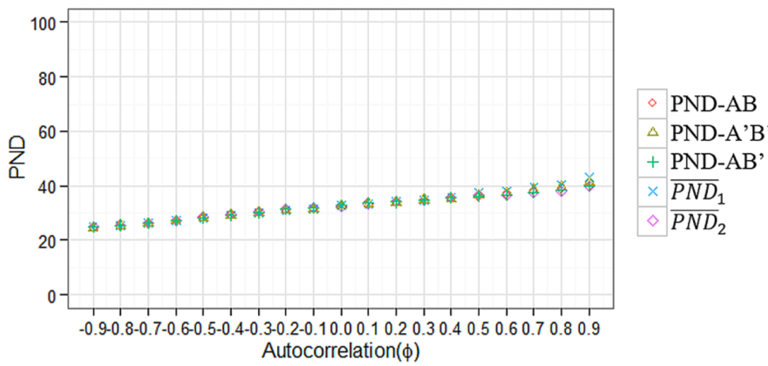
Level and interaction data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)

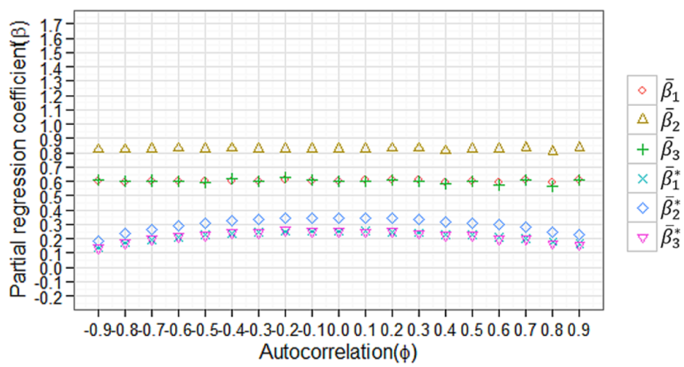






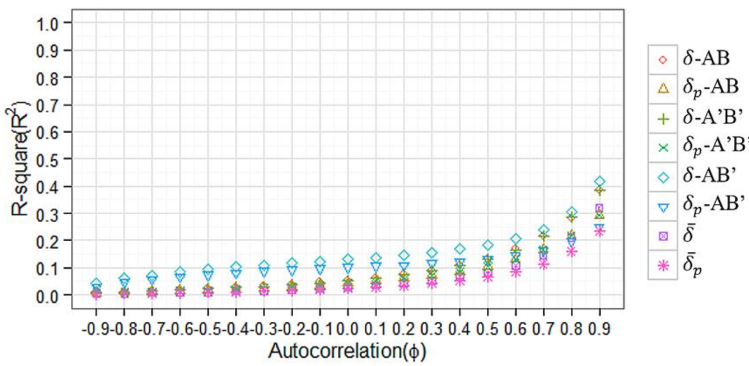

d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

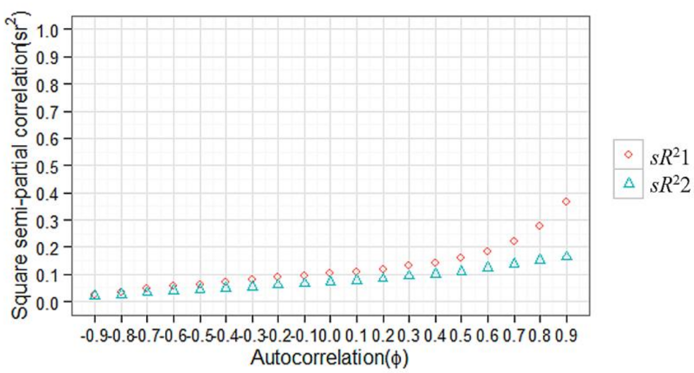

PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

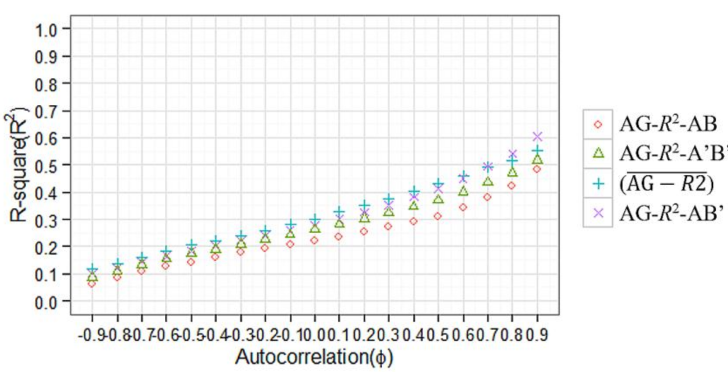

Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

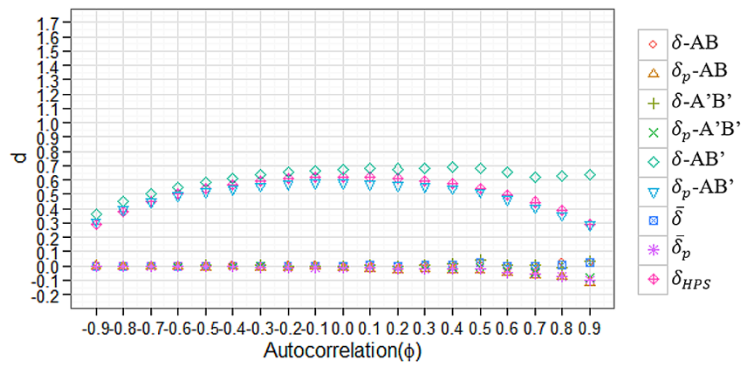
Appendix L

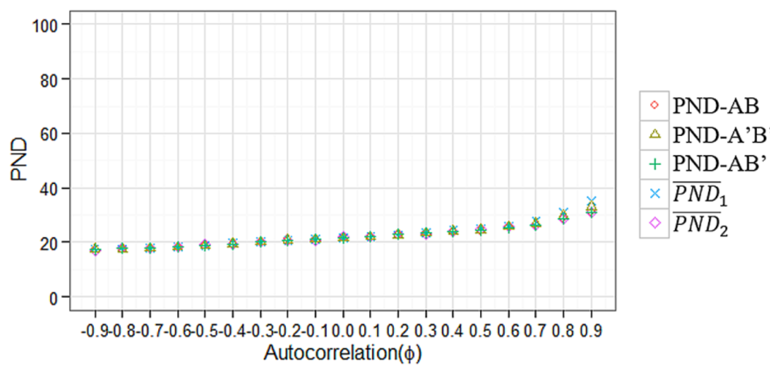
Level and interaction data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

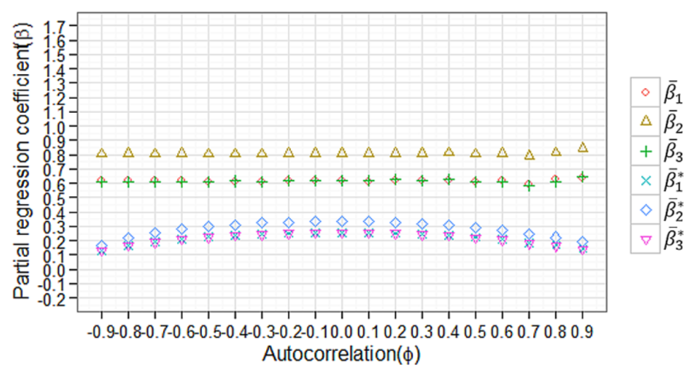






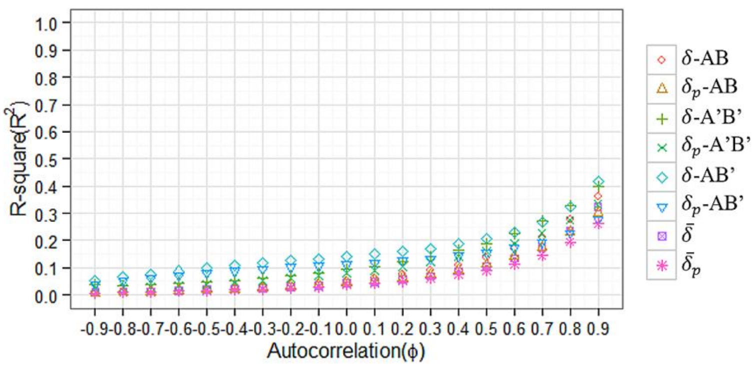

d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

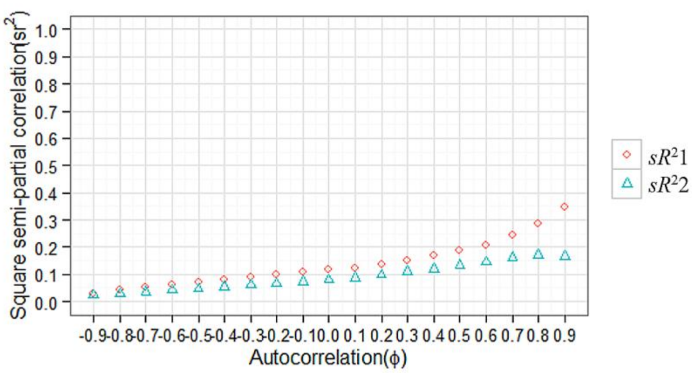

PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

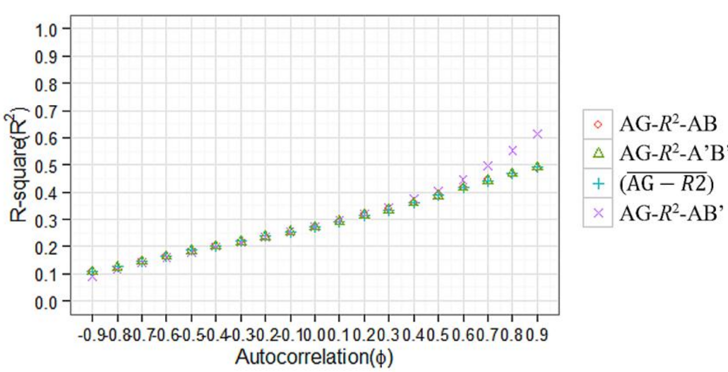

Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

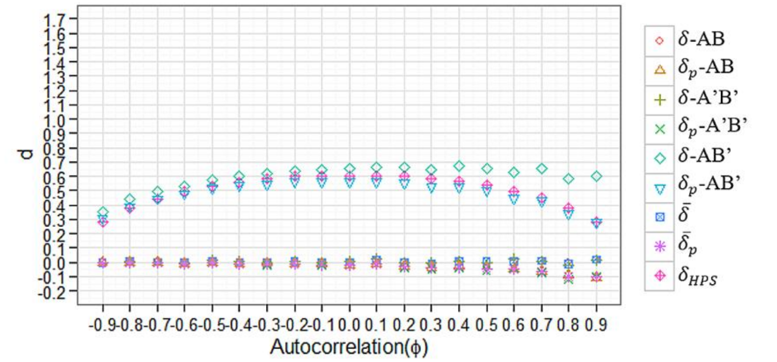
Appendix M

Level and interaction data type, autocorrelation graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)

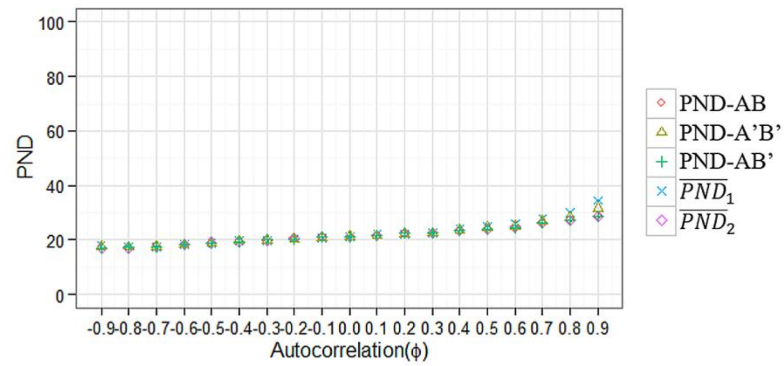




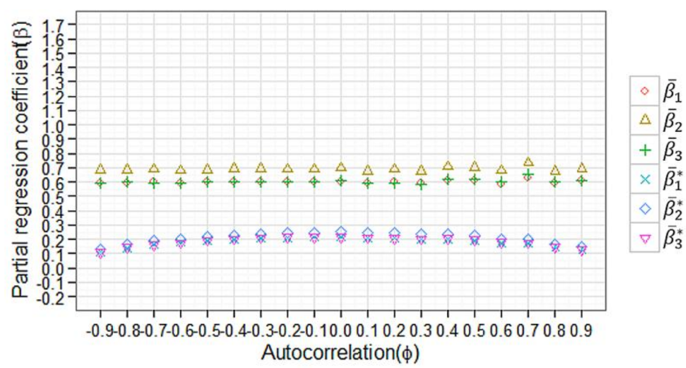




d
Autocorrelation(ϕ)
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}



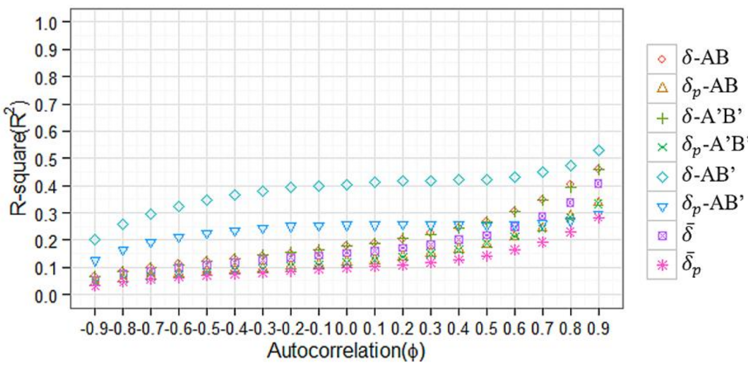
PND
Autocorrelation(ϕ)
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

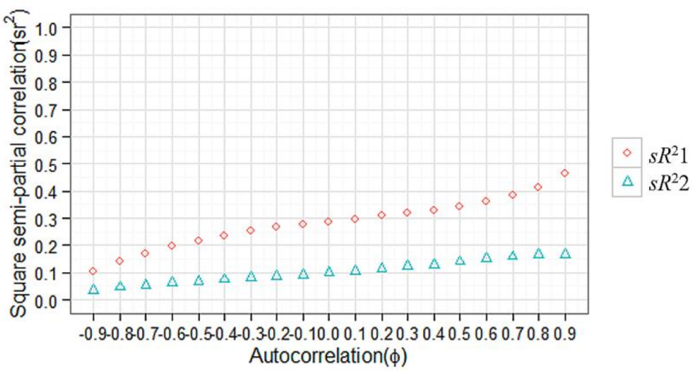


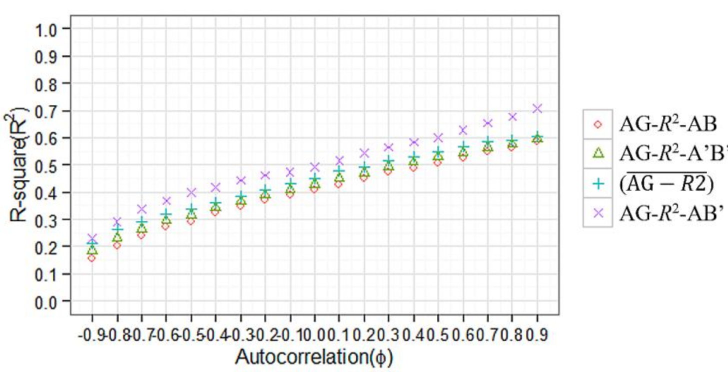
Partial regression coefficient(β)
Autocorrelation(ϕ)
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

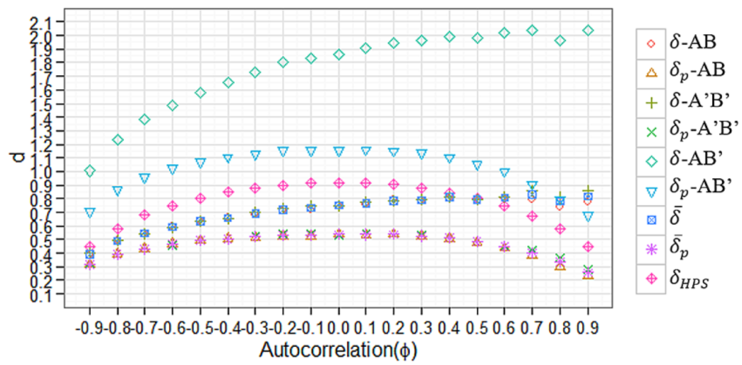
Appendix N

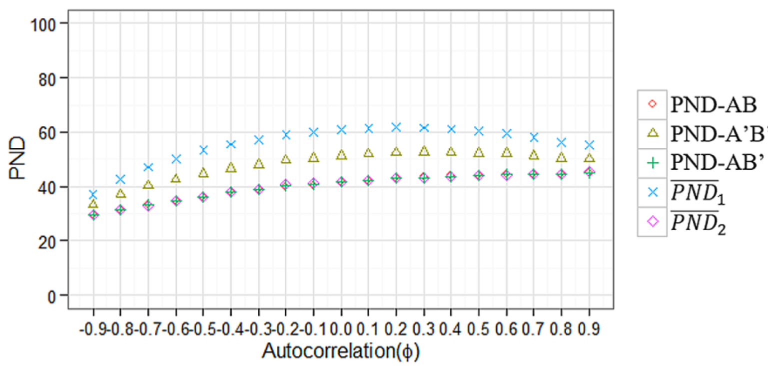
Level, trend and interaction data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)

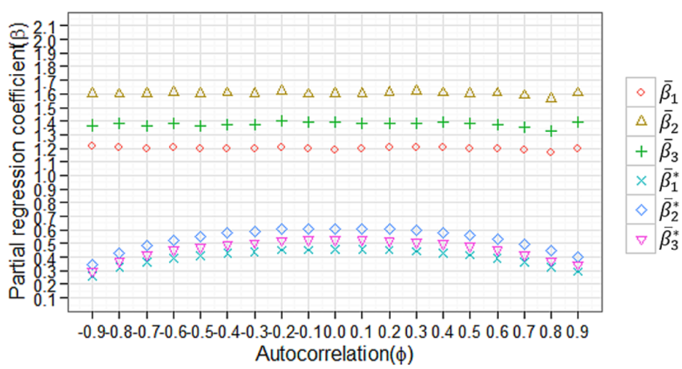






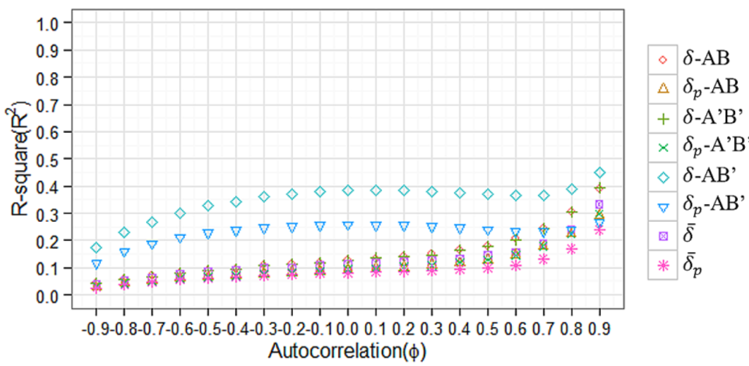

d
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

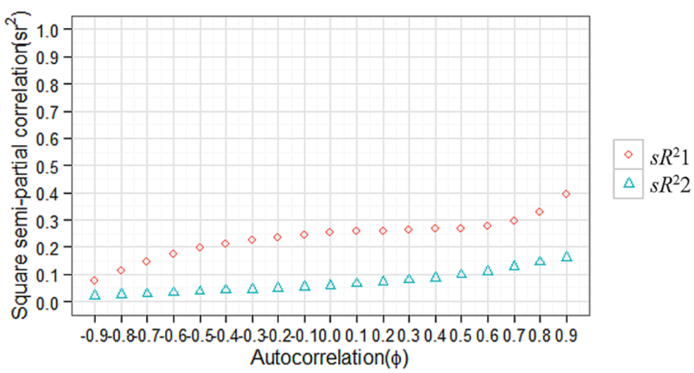

PND
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

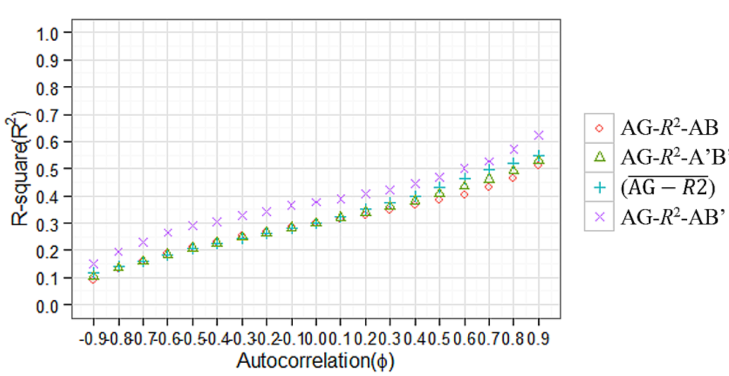

Partial regression coefficient(β)
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

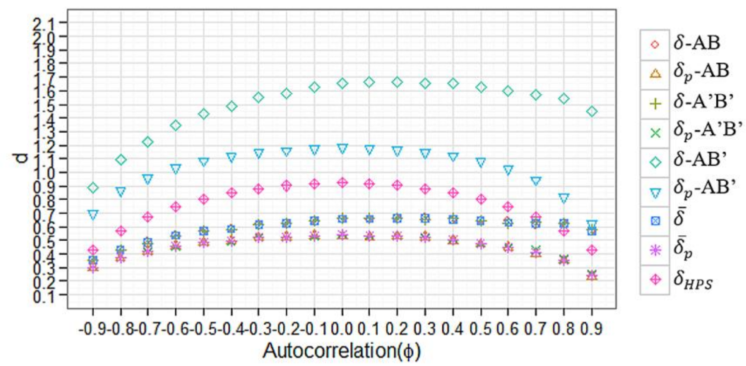
Appendix O

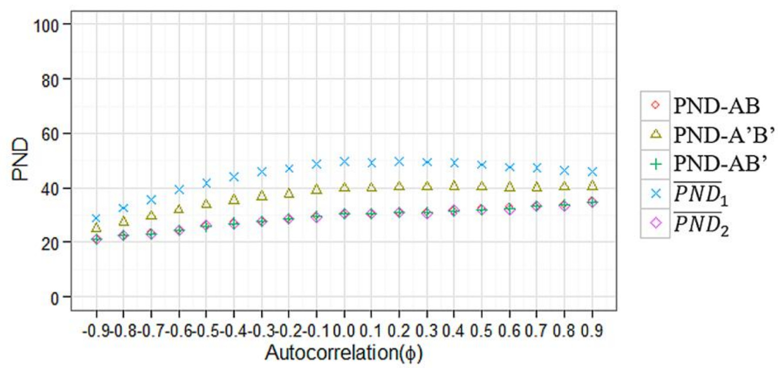
Level, trend and interaction data type, autocorrelation graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

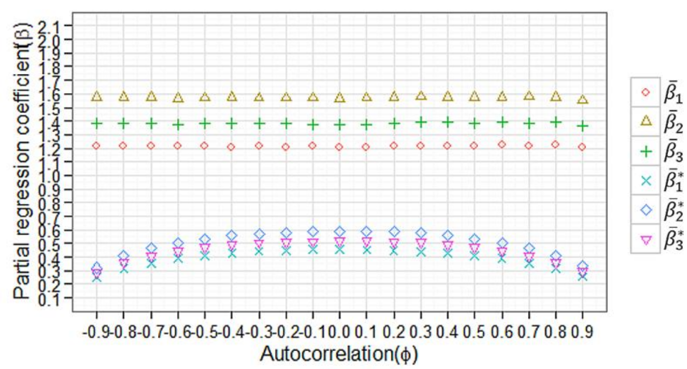






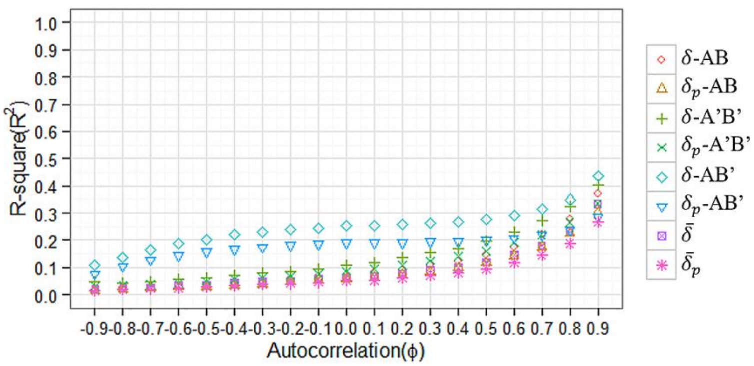

d
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}

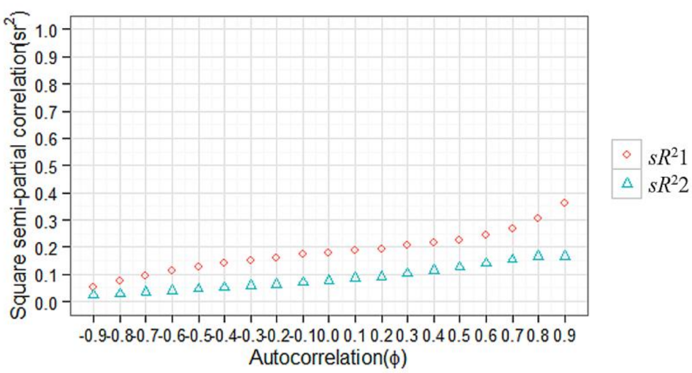

PND
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$

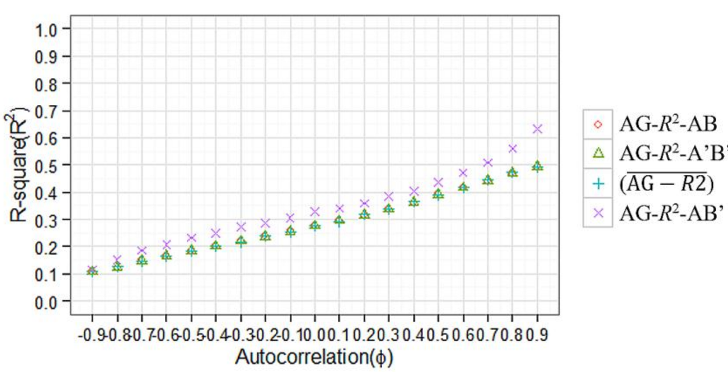

Partial regression coefficient(β)
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

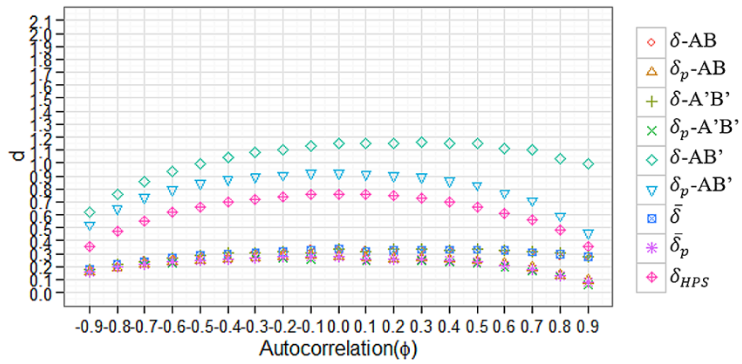
Appendix P

Level, trend and interaction data type, autocorrelation graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)

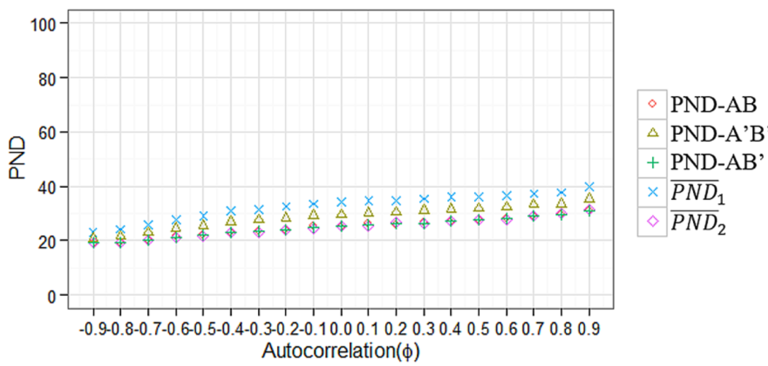




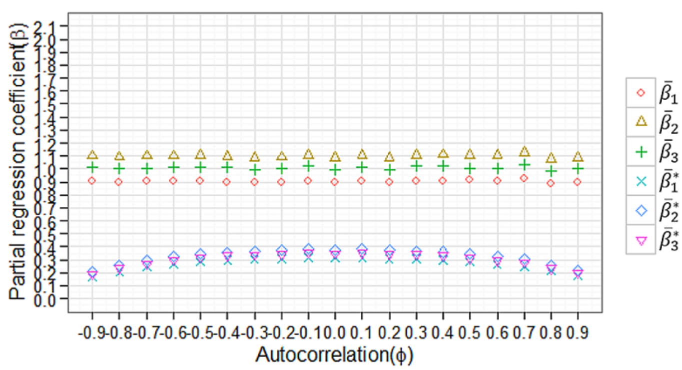




d
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
δ-AB
δ_p-AB
δ-A'B'
δ_p-A'B'
δ-AB'
δ_p-AB'
$\bar{\delta}$
$\bar{\delta}_p$
δ_{HPS}



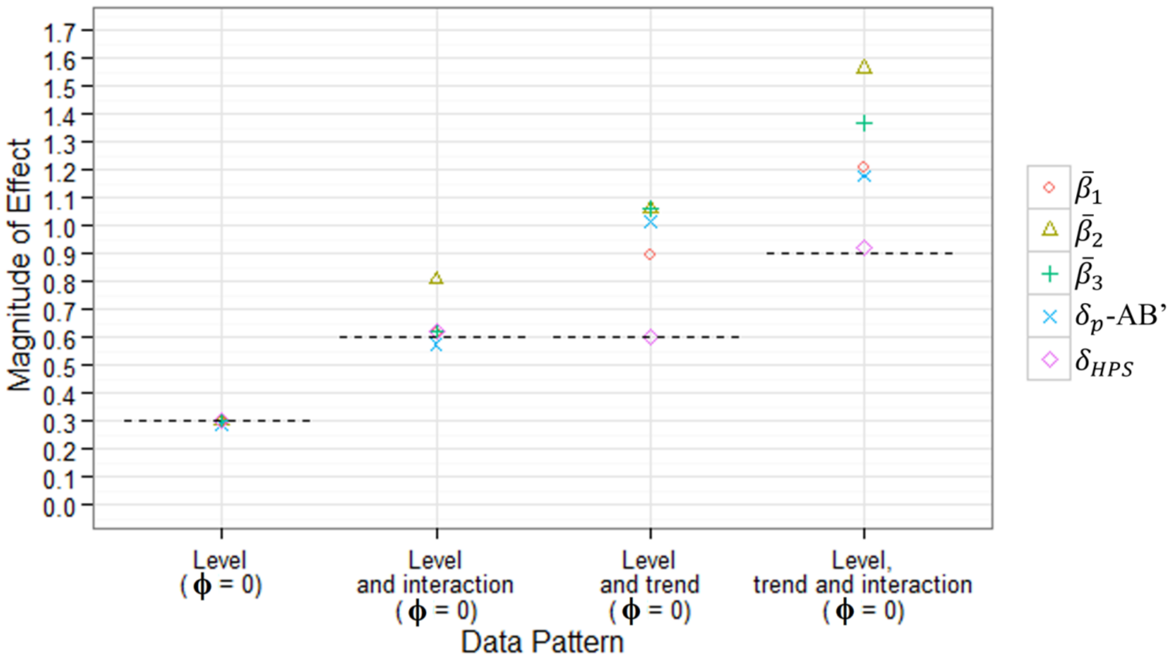
PND
0 20 40 60 80 100
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
PND-AB
PND-A'B'
PND-AB'
$\overline{PND}_1$
$\overline{PND}_2$



Partial regression coefficient(β)
Autocorrelation(ϕ)
-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
$\bar{\beta}_1$
$\bar{\beta}_2$
$\bar{\beta}_3$
$\bar{\beta}_1^*$
$\bar{\beta}_2^*$
$\bar{\beta}_3^*$

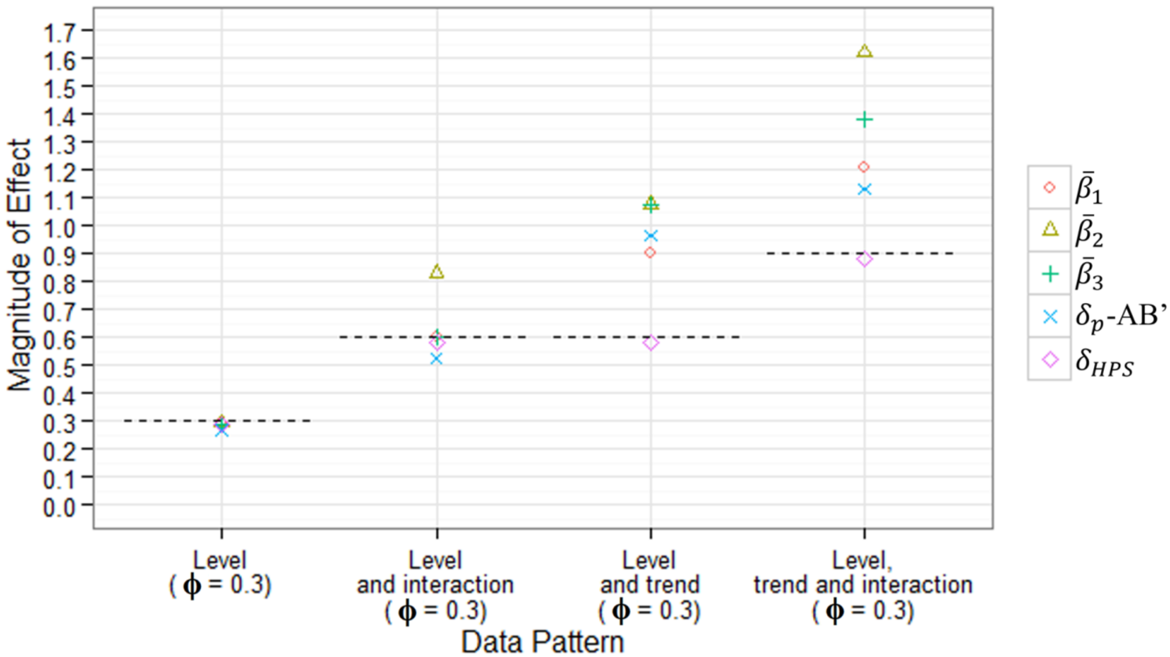
Appendix Q

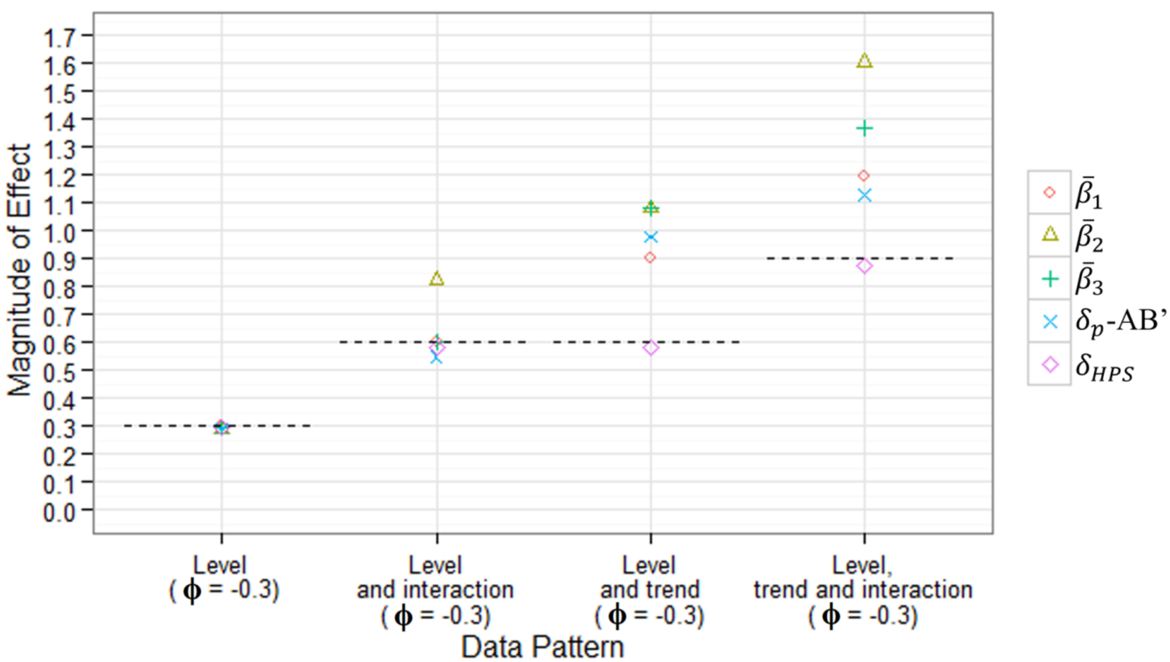
Effects graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)



Appendix R

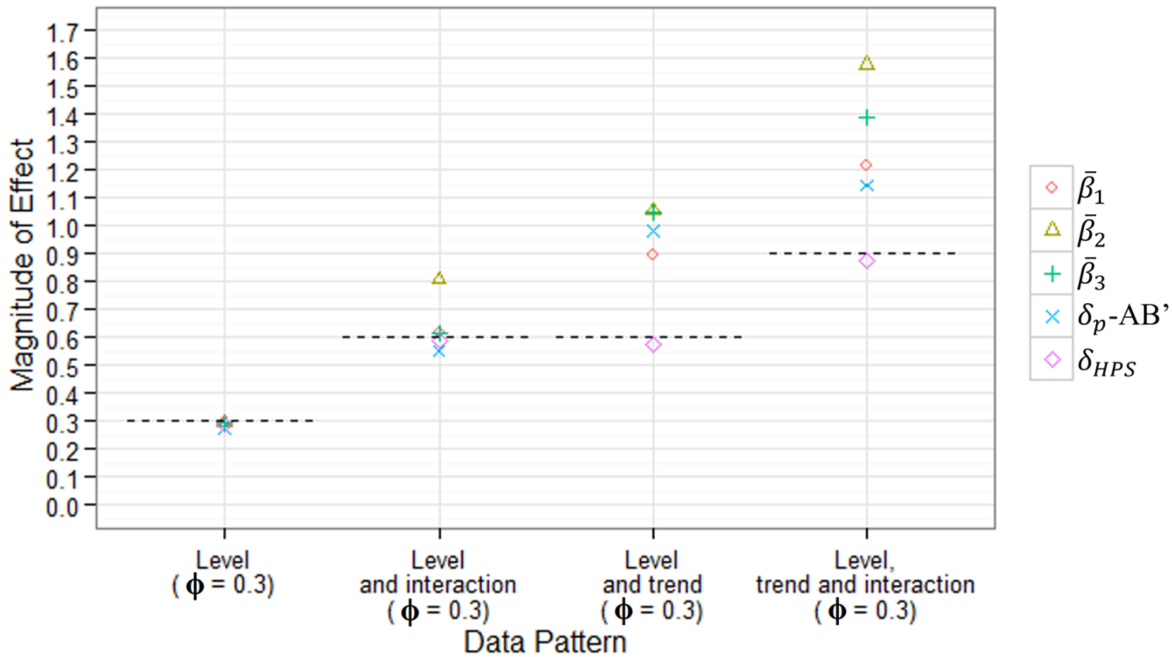
Effects graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)

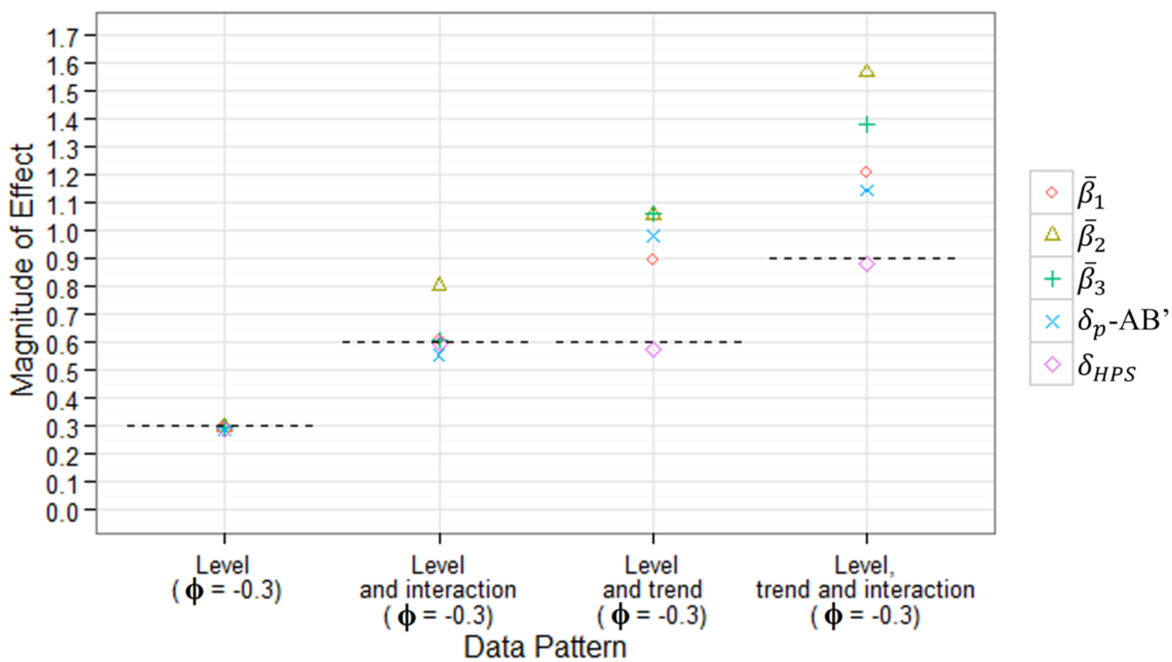




Appendix S

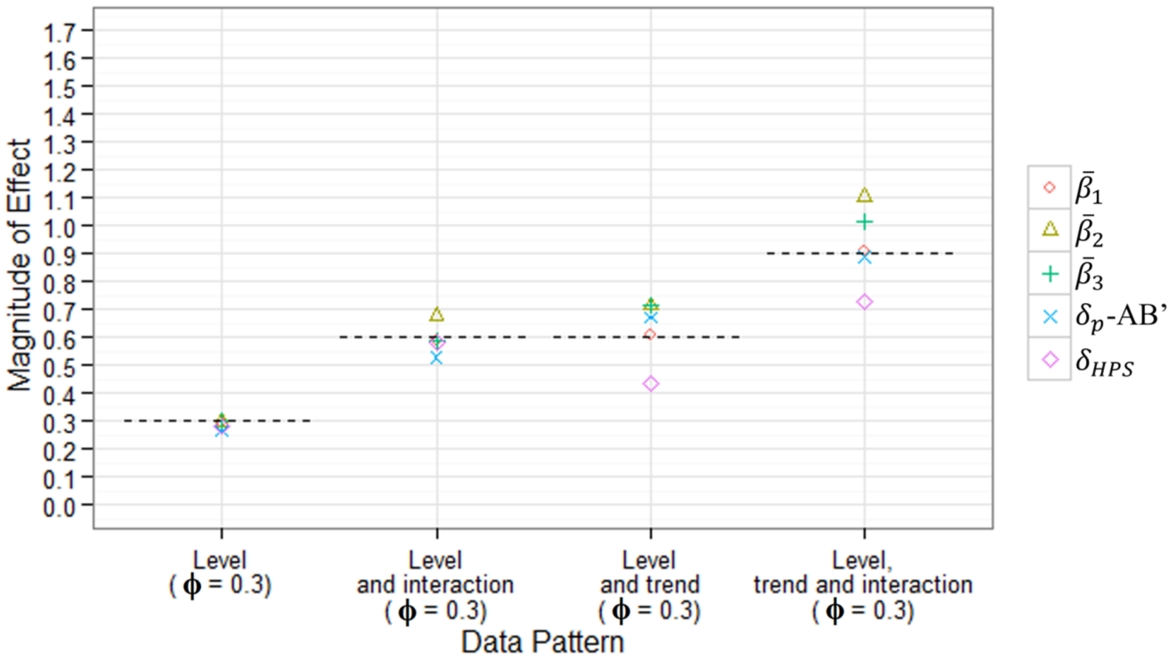
Effects graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)

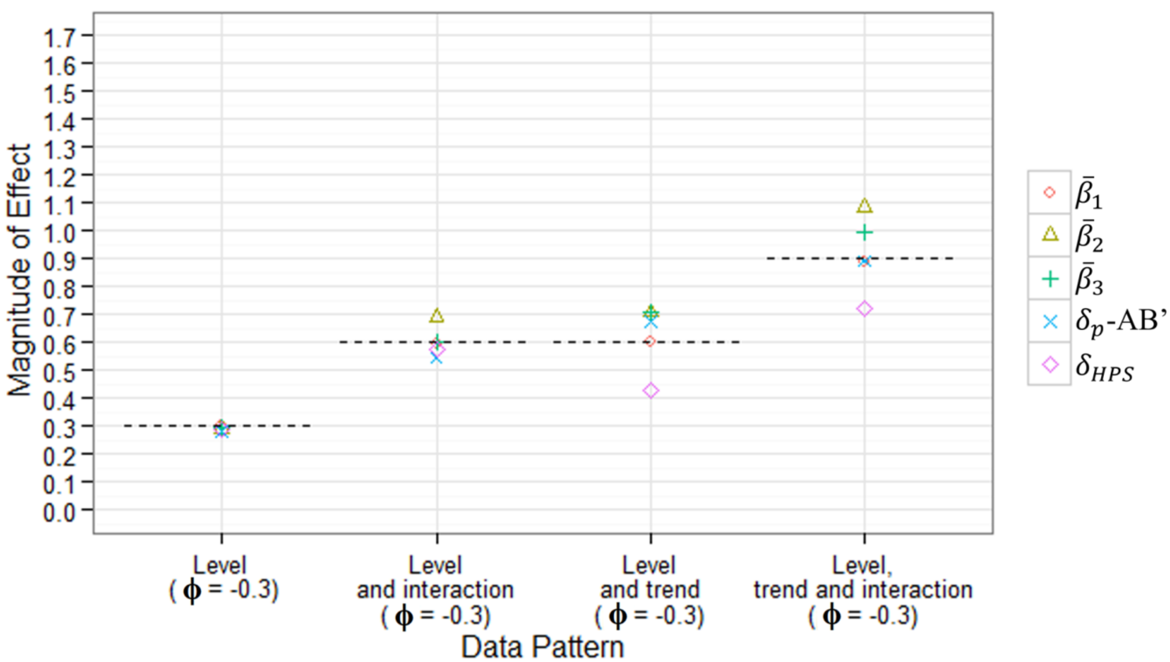




Appendix T

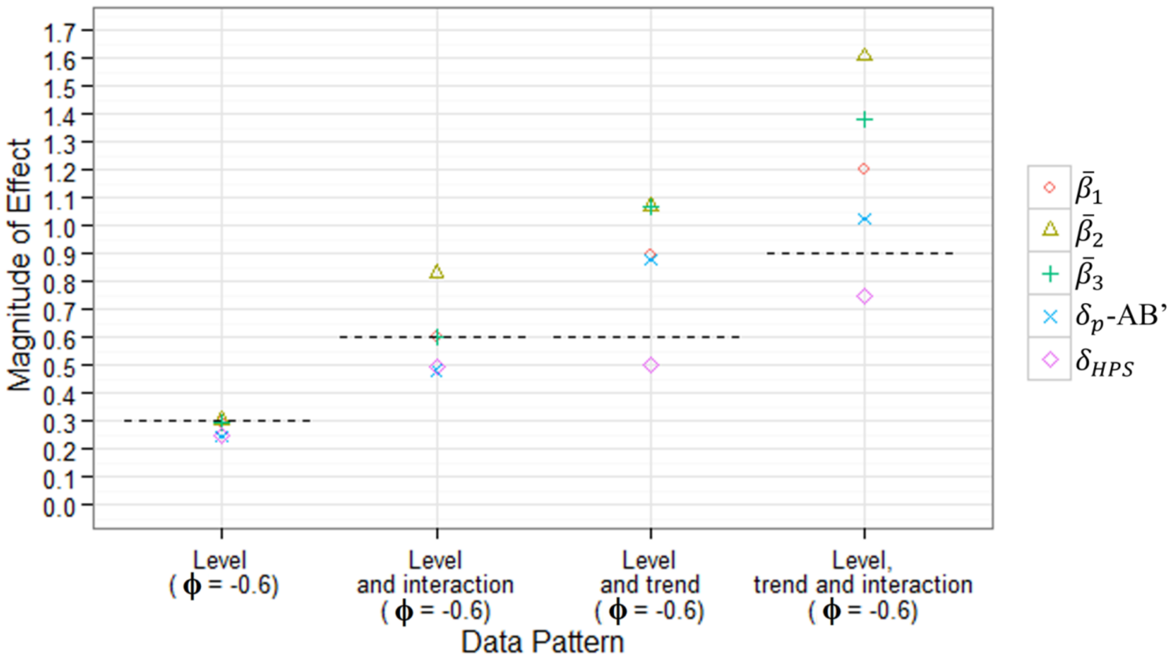
Effects graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)





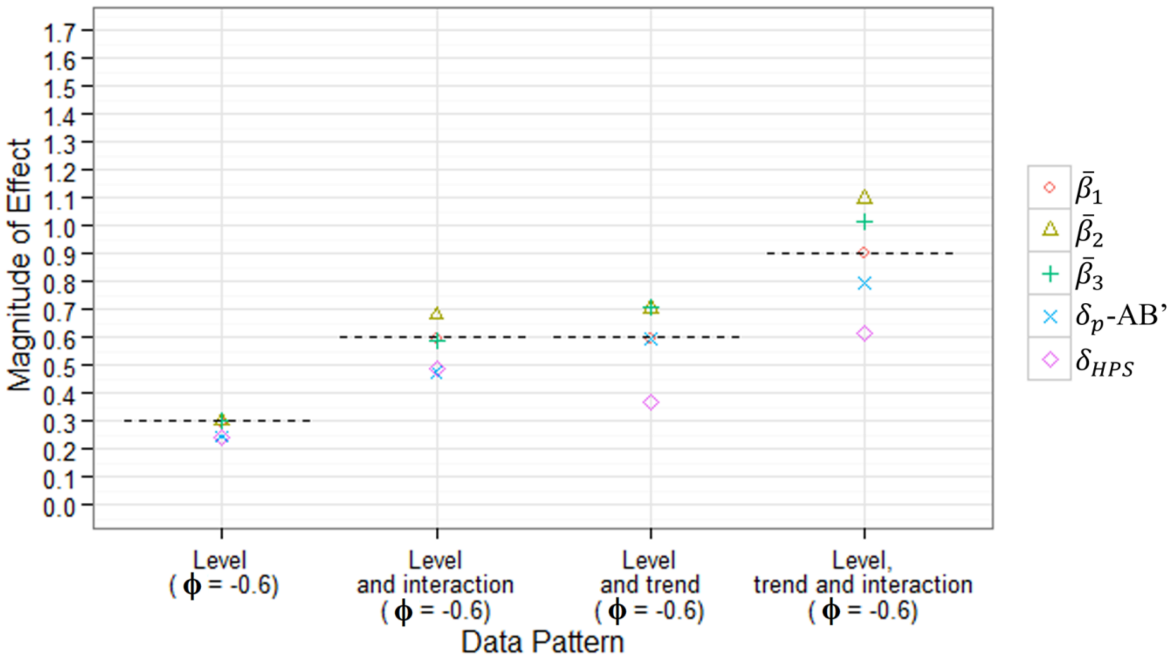
Appendix U

Effects graphs ($t_A = t_B = t_{A'} = t_{B'} = 5$)



Appendix V

Effects graphs ($t_A = t_{A'} = 10,\ t_B = t_{B'} = 5$)



Appendix W

Effects graphs ($t_A = t_B = t_{A'} = t_{B'} = 10$)