TIME SERIES DATA ANALYSIS OF SINGLE SUBJECT EXPERIMENTAL DESIGNS

USING BAYESIAN ESTIMATION

Xing Qin Aerts

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2015

APPROVED:

Prathiba Natesan, Major Professor
Robin Henson, Committee Member
Smita Mehta, Committee Member
Kevin Callahan, Committee Member
Abbas Tashakkori, Chair of the Department of Educational Psychology
Jerry Thomas, Dean of the College of Education
Costas Tsatsoulis, Interim Dean of the Toulouse Graduate School

This study presents a set of data analysis approaches for single subject designs (SSDs). The primary purpose is to establish a series of statistical models to supplement visual analysis in single subject research using Bayesian estimation. Linear modeling approach has been used to study level and trend changes. I propose an alternate approach that treats the phase change-point between the baseline and intervention conditions as an unknown parameter. Similar to some existing approaches, the models take into account changes in slopes and intercepts in the presence of serial dependency. The Bayesian procedure used to estimate the parameters and analyze the data is described. Researchers use a variety of statistical analysis methods to analyze different single subject research designs. This dissertation presents a series of statistical models to model data from various conditions: the baseline phase, A-B design, A-B-A-B design, multiple baseline design, alternating treatments design, and changing criterion design. The change-point evaluation method can provide additional confirmation of causal effect of the treatment on target behavior. Software codes are provided as supplemental materials in the appendices. The applicability for the analyses is demonstrated using five examples from the SSD literature.
Copyright 2015

by

Xing Qin Aerts
# TABLE OF CONTENTS


<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAYESIAN ESTIMATION OF SINGLE SUBJECT DESIGNS: DATA ANALYSIS FOR THE BASELINE, A-B DESIGN, AND A-B-A-B DESIGN .................................................. 1</td>
</tr>
<tr>
<td>Time Series with Serial Dependency ................................................................. 4</td>
</tr>
<tr>
<td>Bayesian Data Analysis ......................................................................................... 5</td>
</tr>
<tr>
<td>Methods ................................................................................................................. 7</td>
</tr>
<tr>
<td>Model Estimation .................................................................................................. 7</td>
</tr>
<tr>
<td>Baseline Data Analysis ....................................................................................... 8</td>
</tr>
<tr>
<td>A-B Design Data Analysis .................................................................................. 11</td>
</tr>
<tr>
<td>Data Analyses for A-B-A-B Designs ................................................................. 14</td>
</tr>
<tr>
<td>Examples .............................................................................................................. 17</td>
</tr>
<tr>
<td>Baseline Data Analysis ....................................................................................... 17</td>
</tr>
<tr>
<td>Data Analysis of the A-B Design ....................................................................... 24</td>
</tr>
<tr>
<td>Data Analysis for A-B-A-B Designs .................................................................. 28</td>
</tr>
<tr>
<td>Discussion .......................................................................................................... 34</td>
</tr>
<tr>
<td>References .......................................................................................................... 38</td>
</tr>
<tr>
<td>Appendix A R and JAGS Codes for the Baseline Data Analysis ......................... 41</td>
</tr>
<tr>
<td>Appendix B R and JAGS Codes for the A-B Design Data Analysis ....................... 44</td>
</tr>
<tr>
<td>Appendix C R and JAGS Codes for the A-B-A-B Design Data Analysis ............... 48</td>
</tr>
</tbody>
</table>

**BAYESIAN ANALYSIS OF MULTIPLE BASELINE DESIGN, ALTERNATING TREATMENTS DESIGN, AND CHANGING CRITERION DESIGN**

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAYESIAN ANALYSIS OF MULTIPLE BASELINE DESIGN, ALTERNATING TREATMENTS DESIGN, AND CHANGING CRITERION DESIGN ........................................... 53</td>
</tr>
<tr>
<td>Methods ................................................................................................................. 54</td>
</tr>
<tr>
<td>Multiple Baseline Design ..................................................................................... 55</td>
</tr>
<tr>
<td>Alternating Treatments Design (ATD) .................................................................. 56</td>
</tr>
</tbody>
</table>
Statistical analysis is one of the primary methods in the data analysis of single subject research (SSR). It is recommended in addition to visual analysis for analyzing SSR data (Gast, 2010; Houle, 2009). Statistical methods provide quantification of the data which is considered by some in the field to be more objective (Campbell & Herzinger, 2010). They can be used to identify smaller effects and help increase confidence in visual analysis (Gorman & Allison, 1997). Table 1 summarizes the strengths and weaknesses of select statistical methods.

Table 1

**Summary of Select Statistical Methods in Single Subject Research**

<table>
<thead>
<tr>
<th>Method</th>
<th>Applicable designs</th>
<th>Description</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>t</em> and <em>F</em> Tests</td>
<td>A-B, A-B-A-B, and multiple baseline designs.</td>
<td>Statistical significance tests on the mean differences in data between phases.</td>
<td>Account for level changes between phases.</td>
<td>Can only be used for data without serial dependency; do not account for trend changes.</td>
</tr>
<tr>
<td>Randomization tests</td>
<td>Withdrawal and reversal designs in which treatment can be implemented and withdrawn repeatedly.</td>
<td>A nonparametric test. Experimental conditions randomly assigned to occasions over time. Calculate the probability of obtaining a difference between the two or more conditions as large as the difference obtained in the study.</td>
<td>Assumptions are less rigorous than parametric statistics.</td>
<td>Cannot be used for irreversible behaviors.</td>
</tr>
<tr>
<td>Busk and Serlin’s Effect size</td>
<td>A-B design.</td>
<td>Include three measurement approaches with the general formula of the mean difference across phases.</td>
<td>Account for level changes, yield to an effect size estimate.</td>
<td>Do not capture trend changes; can only be used for data without serial dependency.</td>
</tr>
</tbody>
</table>

*(table continues)*
Table 1 (continued).

<table>
<thead>
<tr>
<th>Method</th>
<th>Applicable designs</th>
<th>Description</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of non-overlapping data points (PND)</td>
<td>Any single subject design with two or more phases.</td>
<td>The percentage of data points that do not overlap with data across adjacent phases.</td>
<td>Alignment with experimental effects logic. Applicable to any single subject design with two or more phases.</td>
<td>Do not account for variability and trend changes.</td>
</tr>
<tr>
<td>Percent of all non-overlapping data (PAND)</td>
<td>A reversal or Multiple baseline or other longer designs.</td>
<td>Create a 2 by 2 table with all the data points from baseline and intervention phases; calculate Pearson's Phi coefficient.</td>
<td>Uses all the data; effect size measure presented in standard deviation units.</td>
<td>Not recommended for short, single-baseline designs of less than 20 or 25 data points; do not account for within-phase data trends and trend changes.</td>
</tr>
<tr>
<td>Time-series data analysis</td>
<td>Designs with change across phases.</td>
<td>Detect the statistical significance of the level and/or trend change across phases, taken into account autocorrelation.</td>
<td>Suitable for single subject data show serial dependency.</td>
<td>Requires large amount of data points or observations (e.g., 50).</td>
</tr>
<tr>
<td>Hierarchical linear modeling</td>
<td>Designs with multiple subjects.</td>
<td>Estimate individual growth across phases and allow the growth to vary across subjects.</td>
<td>Produces an estimate of treatment effect under multilevel modeling.</td>
<td>Requires large amount of data points or observations for better estimation.</td>
</tr>
<tr>
<td>de Vries and Morey Bayes factors (2014)</td>
<td>2-phases designs.</td>
<td>The ratio of the probability favoring the hypothesis that the intervention effect exists.</td>
<td>Using Bayesian estimation; account for trend and/or level changes; take into account autocorrelation.</td>
<td>Can only be applied to 2-phases designs or the adjacent two phases of other designs.</td>
</tr>
</tbody>
</table>

The primary purpose of the current study is to establish a series of statistical models to supplement visual analysis in SSR using Bayesian estimation. In the present study, I treat the phase change-points between the baseline and intervention conditions as unknown parameters that will be estimated using the Bayesian framework. This can provide additional confirmation for causal effect of the intervention on target behavior. If
the estimated values of the change-point correspond to change in treatment conditions, an intervention effect is indicated. Change in level and trend across phases can be used to quantify the strength or magnitude of the treatment effect.

We develop and demonstrate three time series models to analyze baseline data, the A-B design, and the A-B-A-B design. Due to the time series nature of SSR data, it is necessary to consider serial dependency in the statistical models (Kazdin, 1984; Manolov, Solanas, Sierra, & Evans, 2011; Swaminathan et al., 2014). The current study seeks to answer the following research questions:

(1) How can Bayesian estimation be used to examine the stability and trend of the baseline data with serial dependency in single subject designs (SSD)?

(2) How effectively can Bayesian estimation detect the change-point(s) between phases when the statistical analyst is blind to phase change points?

(3) How can Bayesian estimation of change in the intercept and slope parameters across phases in A-B and A-B-A-B designs be used to quantify the strength or magnitude of intervention?

The models developed in this study are confirmatory models. They are not used for prediction purposes. We seek to provide additional assurance for researchers in interpreting the results accurately. In order to demonstrate the effectiveness of the new methods, we apply the statistical models to the data from published empirical SSR literature.

SSR data contain repeated measures or continuous observations and consequently, have the issue of serial dependency (Kazdin, 1984). Ignoring the existence of serial dependency in statistical analyses of data may lead to
misinterpretation of the intervention effect. To address this issue, SSR data analyses procedures need to incorporate autocorrelation in the statistical model (Kazdin, 1984; Manolov et al., 2011; Swaminathan et al., 2014). In the following section, we are going to discuss the issue of autocorrelation and serial dependency in time series analysis.

**Time Series with Serial Dependency**

Autocorrelation is an important concept in interrupted time series analysis (ITSA) (Kazdin, 1984; McDowall, McCleary, Meidinger, & Hay, 1980). Autocorrelation refers to the correlation between the current data points and the previous data points (i.e., lags). However, independence of errors is a critical assumption in statistical significance tests (SST, Kazdin, 1984). The expected value of the error correlation is assumed to be zero ($r_{e_1,e_2} = 0$) in SST. In time series data with repeated measures, the assumption of independence of errors is usually not met. Serial dependency occurs in this type of data with repeated measures when the autocorrelation of error terms does not equal to zero (Kazdin, 1984).

In ITSA, autocorrelation parameter indicates the order of the autocorrelation. For example, if the error is only correlated with its first lag (lag-1), the autocorrelation parameter is one; if the error is also correlated with its second lag (lag-2), the autocorrelation parameter is two. In the social sciences, the time series are usually well represented by lower order autocorrelation. The autocorrelation parameter rarely exceeds one (McDowall et al., 1980). Autocorrelation of lag-1 indicates serial dependency in SSR data (Kazdin, 1984). Swaminathan et al. (2014) presented the equations of the linear regression model and the serial dependency of the residual ($e_t$) as,
In equations 1 and 2, $y_t$ is the predicted value of the target behavior at time $t$; $\beta_0$ and $\beta_1$ are the intercept and slope coefficients of the linear regression model, respectively; $e_t$ is the error at time $t$; $\rho$ is the autocorrelation coefficient; and $\epsilon_t$ is the independently distributed error. In ITSA, this error is referred to as the white noise (McDowall et al., 1980). Combining these two equations and transforming, the regression model with serial dependency becomes:

$y_t - \rho y_{t-1} = \beta_0 (1 - \rho) + \beta_1 [t - \rho (t - 1)] + \epsilon_t$  \hspace{1cm} (3)

In the present study, equation 3 serves as the basic model for the time series data analysis. The specific models for different types of analyses are demonstrated in detail in the method section.

Bayesian Data Analysis

Bayesian methods maximize the utility of the sample data. The posterior distribution is updated iteratively until the parameter estimates converge (Kruschke, 2011). The highest density interval (HDI) of the posterior distribution represents the parameter estimates with the highest occurrence in the iteration process. In SSR data analysis, accurately estimating the parameter of the data pattern change indicates replication of the effect of the intervention. We provide more detailed explanation in the section of the A-B design data analysis example.

---

For those interested in greater detail:

$y_t = \beta_0 + \beta_1 t + e_t$

$e_t = \rho e_{t-1} + \epsilon_t$  \hspace{1cm} (2)

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$

$y_t = \beta_0 + \beta_1 t + e_t$
Compared with maximum likelihood-based approaches, Bayesian probability theory has some advantages. First of all, Bayesian framework gives probabilities to all the possible values of the unknown parameter. It expands the single valued hypothesis into the prior distribution (Kruschke, 2011; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). For example, consider the mean difference as the parameter of interest ($\theta$). The inference in statistical significance tests (SST) is based on the assumption that the null hypothesis is true, or $p(H_0: \theta = 0) = 1$. The inferential in Bayesian framework is based on the prior belief, or $\theta \sim p(\theta)$.

Moreover, Bayesian framework examines the probability of the unknown parameter [$p(\theta|\text{Data})$], instead of the probability of the sample data [$p(\text{Data}|\theta)$]. SST is based on the significance level of $p(\text{Data}|\theta)$ (the probability of obtaining the sample data given the null hypothesis is true, or the likelihood; Kruschke, 2011). Bayesian theory applies the likelihood onto the prior distribution of the parameter [$p(\theta)$] and updates the prior belief into the posterior distribution [$p(\theta|\text{Data})$], the probability of the estimated parameter given the observed sample data:

$$p(\theta|\text{Data}) \propto p(\text{Data}|\theta) \times p(\theta)\quad (4)$$

Bayesian estimation methods have already been applied to SSR data analyses. Zucker, Schmid, McIntosh, D'Agostino, Selker, and Lau (1997) applied a hierarchical Bayesian random effect model to combine single patient studies under the randomized controlled multi-crossover design. They applied the structured design to assess the effects of medical treatment on chronic diseases. Swaminathan et al. (2014) applied Bayesian method in a multilevel model that accounted for autocorrelation of error terms. de Vries and Morey (2013) introduced a set of Bayes factor tests for single subject data.
with autocorrelation. Swaminathan’s model is applicable for the A-B and A-B-A-B designs with multiple subjects. Whereas de Vries and Morey’s Bayes factors can be applied for the two phases design (e.g., A-B) or the adjacent two phases of other designs (e.g. A-B and B-A in A-B-A-B). However, none of these studies have presented a comprehensive approach to evaluate the quality of parameter estimates by analyzing the posterior distribution. Nor have these studies evaluated intervention effect by considering the change-point as an unknown parameter. The present study seeks to fill this gap. The approach presented in the current study considers the change-point as an unknown parameter. Examining the posterior of the change-point as an estimate of change in the treatment provides additional confirmation of the intervention effect. Examining the change in level and trend of the baseline and intervention phases quantifies the magnitude of the treatment. This is the application of Bayesian test of differences in SSDs.

Methods

The current study introduces Bayesian data analysis methods for SSR designs. The basic model formulas were presented in equations 1, 2, and 3 above. For different types of SSR designs, the application of the specific models needs to be different. Model Estimation

The present study uses the software program JAGS 3.4.0 (Plummer, 2003) to demonstrate Bayesian estimation of different designs in SSR. Data analysis is executed through R 3.1.0 (R Core Team, 2014). This process can be applied using the R packages rjags and runjags. The rjags package is used to run a user-specified JAGS model from within R whereas the runjags package needs a method such as rjags to
transfer the computation from R to JAGS. The autorun.jags command in runjags calculates the run length and diagnoses convergence automatically. runjags uses the Gelman and Rubin statistic (Gelman & Rubin, 1992) to diagnose convergence. The potential scale reduction factor should be below 1.05 for all parameters. The output of autorun.jags provides the posterior distributions of the parameters.

The shape of the posterior distribution indicates the accuracy of the parameter estimates. The most prominent mode indicates the most probable value of the change-point. The mean, standard deviation, and the 95% highest density interval (HDI) in the output summary are used to describe the intercept (level), slope (trend), change in intercept and slope across phases, autocorrelation (serial dependency), and standard deviation (stability). The 95% HDI can be found in the output summary as the “Lower95” and “Upper95” of the parameter estimate. The estimation criterion to reject the null hypothesis is that if the parameter’s region of practical equivalence (ROPE) falls outside its 95% HDI of the posterior distribution (Kruschke, 2011). ROPE is a small range of values that are equivalent to the value of interest for all practical purposes. It needs to be determined before the analysis is conducted.

Baseline Data Analysis

In SSR, baseline data are recorded and plotted for every time point (Gast & Spriggs, 2010). The intervention phase is implemented once a stable data pattern in the baseline phase is observed. A criterion for stability in visual analysis is to obtain three or more data points with similar values (Kennedy, 2005). Another commonly used criterion for level stability is when 80% of the data fall into the 20-25% range of the median (Gast
However, these popular methods are not always helpful in the SSR practice.

In this section, we introduce a statistical model that can estimate the stability and trend of the baseline data. The application of the baseline data analysis can be expanded to different conditions in other designs. For example, it can be applied to the intervention phase of the A-B design to evaluate the stability of the second phase without the influence of the stability of the first phase. In other words, the baseline data analysis is a within phase data analysis which can be applied to any one phase of SSR designs.

A linear model is established with trend and serial dependency taken into account. In order to include the effect of autocorrelation into the model, the first data point needs to be defined separately from the rest of the time series. This is because the first time point of the time series does not have its lag-1 time point. The statistical model for the first time point is presented in equation 5.

\[
\hat{y}_1 = \text{int} + \beta_1, \quad \text{int} = \text{intercept (the theoretical value of the observation at time point zero)}; \quad \beta_1 = \text{the slope.}
\]

Let us assume that the observed value at the first time point \(y_1\) follows a normal distribution with the mean of \(\hat{y}_1\) and standard deviation of \(\sigma_e\). Therefore, the precision of the observation \(y_1\) is \(\tau_e\).

\[
y_1 \sim \text{dnorm}(\hat{y}_1, \tau_e), \quad \tau_e = \frac{1}{\sigma_e^2}
\]

In equation 5, \(\hat{y}_1\) is the predicted value of the first time point; \(\text{int}\) is the intercept (the theoretical value of the observation at time point zero); \(\beta_1\) is the slope. Let us assume that the observed value at the first time point \(y_1\) follows a normal distribution with the mean of \(\hat{y}_1\) and standard deviation of \(\sigma_e\). Therefore, the precision of the observation \(y_1\) is \(\tau_e\).

The rest of the time series follow a linear procedure with 1-lag autocorrelated errors. According to the linear function (equation 1), the error autocorrelation (equation
2), and the combined formula after transformation (equation 3), the predicted value for 
\( y_t \) \((t \geq 2)\) is shown in equation 6.

\[
\hat{y}_t = \text{int}(1 - \rho) + \beta_1[t - \rho(t - 1)] + \rho y_{t-1}, \quad (t \geq 2)
\]

\[
y_t \sim \text{dnorm}(\hat{y}_t, \tau_d), \quad \tau_d = \frac{1}{\sigma_e^2}
\]

Equation 7 presents the relation between \( \sigma_e \) and \( \sigma_e^2 \) (Swaminathan et al., 2014).

Priors for the estimated parameters are specified in equations 8, 9, 10, and 11. The 
prior distribution for the intercept in equation 8 reflects the range of the dependent 
variable \([\min(y) - \max(y)]\).

\[
\beta_1 \sim \text{dnorm}(0, \text{prec}_{s_1}), \quad \text{prec}_{s_1} \sim \text{dgamma}(1, 1)
\]

\[
\rho \sim \text{dbeta}(1, 5)
\]

\[
\tau_d \sim \text{dgamma}(1, 1)
\]

In the statistical model (equations 5 and 6), the stability of the data is reflected by 
the variance of the observations \( (\sigma_e^2) \). When the estimated standard deviation (i.e.,

---

\[2 \text{ In more detail:}
\]

\[e_t = \rho e_{t-1} + \epsilon_t, \quad \Rightarrow \epsilon_t = e_t - \rho e_{t-1}\]

The variance of \( \epsilon_t \) is:

\[
\sigma_e^2 = \frac{\Sigma (e_t - \rho e_{t-1} - \epsilon_t)^2}{t}
\]

\[
= \frac{1}{t}[e_t^2 + (e_t^2 - 2\rho e_t e_{t-1} + \rho^2 e_{t-1}^2) + \cdots + (e_t^2 - 2\rho e_t e_{t-1} + \rho^2 e_{t-1}^2)]
\]

\[
= \frac{1}{t}[(e_t^2 + e_{t-1}^2 + \cdots + e_{t-1}^2) - 2\rho(e_t e_{t-1} + e_{t-1} e_{t-2} + \cdots + e_{t-1} e_{t-1}) + \rho^2(e_t^2 + e_{t-1}^2 + \cdots + e_{t-1}^2)]
\]

\[
= \frac{\Sigma e_t^2 - 2\rho \Sigma e_t e_{t-1} + \rho^2 \Sigma e_{t-1}^2}{t}
\]

\[
= \sigma_{e_t}^2 - 2\rho \text{cov}(e_t, e_{t-1}) + \rho^2 \sigma_{e_{t-1}}^2
\]

Since \( \sigma_{e_t}^2 = \sigma_{e_{t-1}}^2 = \sigma_e^2 \), and by definition \( \rho = \frac{\text{cov}(e_t, e_{t-1})}{\sigma_e^2} \), or \( \text{cov}(e_t, e_{t-1}) = \rho \sigma_e^2 \).

\[
\sigma_e^2 = \sigma_{e_t}^2 - 2\rho \sigma_e^2 + \rho^2 \sigma_e^2 = (1 - \rho^2)\sigma_e^2
\]

Therefore, the relation between \( \sigma_e \) and \( \sigma_e^2 \) is: \( \sigma_e^2 = (1 - \rho^2)\sigma_e^2 \), and equation 10 was true: \( \sigma_e = \frac{\sigma_e}{\sqrt{1 - \rho^2}} \)
stability level) shows minimal variability across different segments of baseline, the stability of baseline data is established. The trend of the baseline data is reflected by the slope ($\beta_1$). In the baseline phase, when the data show a counterproductive trend for the target behavior, the intervention should not be implemented. The parameters are estimated using Markov chain Monte Carlo (MCMC) methods. R and JAGS codes can be found in Appendix A.

Statistical estimation should always coordinate with visual analysis. For example, if the baseline data do not follow a linear function according to the scatter plot of the time series, equations 5 and 6 should be replaced by appropriate polynomial functions. In this paper we focus on linear models.

A-B Design Data Analysis

The A-B design is the most basic design in SSR even though it is not said to be experimentally rigorous by itself. It features one baseline (pre-intervention) phase and one intervention phase. The number of data points in each phase could range from 3 to 10 or more. The A-B design examines the intervention effect based on within subject comparison between baseline and intervention phases. The same comparison is also found in any other design that has a baseline followed by an intervention phase. The A-B design is the building block for other designs in SSR. For example, the pattern of the A-B design can be found in each tier of the multiple baseline design. In the proposed model, the change-point value ($knot$) is set as an unknown parameter. We compare the posterior distribution of the change-point with its actual value. Detecting change-point between the two phases by fitting the appropriate model to the data, verifies the change
occurred at the correct point in time. The linear function for the A-B design is shown in equation 12.

\[ y_t = (1 - D_t)[(int_1 - \beta_1 \times knot) + \beta_1 t] + D_t[(int_2 - \beta_2 \times knot) + \beta_2 t] + e_t, \quad (12) \]

where \( D_t \) is a dummy variable. In baseline phase, \( D_t = 0 \) (when \( t < knot \)); and in intervention phase, \( D_t = 1 \) (when \( t \geq knot \)).

In order to include the effect of the error autocorrelation in the model, the first data point is defined separately from the rest of the time series. The first time point \( (y_1) \) in the baseline phase follows equation 13.

\[ \hat{y}_1 = (int_1 - \beta_1 \times knot) + \beta_1, \quad (13) \]

\[ y_1 \sim \text{dnorm}(\hat{y}_1, \tau_e), \quad \tau_e = \frac{1}{\sigma_e^2} \]

The rest of the time series across two phases \( (y_t) \) follow a linear procedure with 1-lag autocorrelated errors. \( y_t \) (\( t \geq 2 \)) can be modeled with equation 14. Equation 14 is a combination of the linear function (equation 12) and the error autocorrelation equation (equation 2). The transformation is similar to the procedure explained in note 1.

\[ \hat{y}_t = (int_1 - \beta_1 \times knot)[(1 - D_t) - \rho(1 - D_{t-1})] \]

\[ + \beta_1[(1 - D_t)t - \rho(1 - D_{t-1})(t - 1)] \]

\[ + (int_2 - \beta_2 \times knot)(D_t - \rho D_{t-1}) \]

\[ + \beta_2[D_t(t - \rho D_{t-1}(t - 1)] + \rho y_{t-1}, \quad (t \geq 2) \]

\[ y_t \sim \text{dnorm}(\hat{y}_t, \tau_d), \quad \tau_d = \frac{1}{\sigma_d^2}, \quad \sigma_d = \frac{\sigma_e}{\sqrt{1 - \rho^2}} \]

In equations 12, 13 and 14, \( \beta_1 \) and \( \beta_2 \) represent the slopes of the baseline and intervention phases, respectively. In order to compare the behavior change at the point of actual phase change, \( int_1 \) and \( int_2 \) are defined as the predicted value of the
observation at the change-point according to the linear models in baseline and intervention phases, respectively (Figure 1). Therefore, the real intercepts (predicted values at $t = 0$, or $\hat{y}_0$) for the two linear models are $(\text{int}_1 - \beta_1 \times \text{knot})$ and $(\text{int}_2 - \beta_2 \times \text{knot})$.

![Figure 1. Intercepts for the linear model in the A-B design (equations 12, 13, and 14).](image)

Monitored parameters of the Bayesian estimation include $\text{int}_1$, $\text{int}_2$, and $\text{dint}$, levels in the two phases and the level change ($\text{dint} = \text{int}_2 - \text{int}_1$); $\beta_1$, $\beta_2$ and $\text{dslope}$, trends in the two phases and the slope change ($\text{dslope} = \beta_2 - \beta_1$); $\text{knot}$, the estimated change-point; $\rho$, the autocorrelation; and $\sigma_\epsilon$, the standard deviation of the observation. R and JAGS codes for A-B-A-B design can be found in Appendix B. The R package (BayesSingleSub) for computing de Vries and Morey’s Bayes factors (2013) was used for the A-B design example. R code for Bayes factors (de Vries & Morey, 2013) can be found at the end of Appendix B. In Bayesian data analysis, Bayes factor (BF) is used for model comparison (Kruschke, 2011). Bayes factors are the ratio of the probability that
the null hypothesis is true over the probability that the alternative hypothesis is true. de Vries and Morey (2013) proposed a set of Bayes factors, including Bayes factor for level comparison (B\textsubscript{JZS}); Bayes factor for level comparison with autocorrelation in the model (B\textsubscript{JZS+AR}); and Bayes factors for level and trend comparison with autocorrelation (B\textsubscript{trend}, B\textsubscript{int}, and B\textsubscript{i+t}). These Bayes factors can be applied to the two phases designs, to evaluate the level change under assumption of a flat series (B\textsubscript{JZS+AR}), level change only (B\textsubscript{int}), trend change only (B\textsubscript{trend}), and both level and trend changes (B\textsubscript{i+t}). Smaller Bayes factor indicates more evidence for the alternative hypothesis. Bayes factor values smaller than 1 indicate that the null hypothesis is rejected and the intervention effect exists.

Data Analyses for A-B-A-B Designs

The A-B-A-B design is a type of withdrawal and reversal design. It has become one of the commonly used single subject designs in behavioral science (Gast & Hammond, 2010; Shadish & Sullivan, 2011). There are two occasions when the A-B-A-B design should not be used. First, when the dependent variable is not reversible after the intervention is implemented, an A-B-A-B design is not appropriate. Second, when ethical reasons require the effective intervention to be maintained, it should not be used either. For example, if the purpose of the experiment is to reduce the consumption of cigarettes, it should not be reversed after the participant responding to the intervention.

Following the intervention phase in an A-B design (phase B, or B\textsubscript{1} as called in the A-B-A-B design), the A-B-A-B design adds additional phases to replicate the initial effect. A reversal to the baseline conditioned phase (A\textsubscript{2}) constitutes withdrawal of the intervention. In the last phase (B\textsubscript{2}), the intervention is reinstituted, both of which are
done to provide a documentation of effect. The most important feature of this design is that it provides an evaluation of similarity of the data pattern across similar phases through a replication effect (Gast & Hammond, 2010).

Swaminathan et al. (2014) evaluated an A-B-A-B design with multiple participants by fitting a multilevel linear model. The model used in the present study is similar but not identical to the model in Swaminathan et al. (2014). Their model specified the values of the change-points, while the current model estimates the change-point values as unknown parameters \((knot1, knot2, and knot3)\). The intercepts have been reparameterized but the models are equivalent except for the unknown change-points. The basic model is presented in equation 15. The dummy variables allow the observation \((y_t)\) in different phases to apply different intercepts and slopes.

\[
y_t = (1 - D1_t)(int_1 - \beta_1 \times knot1) + \beta_1 t \\
+ D1_t(1 - D2_t)(int_2 - \beta_2 \times knot1) + \beta_2 t \\
+ D1_t D2_t(1 - D3_t)(int_3 - \beta_3 \times knot3) + \beta_3 t \\
+ D1_t D2_t D3_t(int_4 - \beta_4 \times knot3) + \beta_4 t + e_t
\]  

\(D1_t, D2_t,\) and \(D3_t\) are dummy variables. The value for each dummy variable at each phase is presented in Table 2.

Table 2

Values of Dummy Variables for Each Phase for the A-B-A-B Design

<table>
<thead>
<tr>
<th></th>
<th>Baseline A1</th>
<th>Intervention B1</th>
<th>Baseline A2</th>
<th>Intervention B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 \leq t &lt; knot1)</td>
<td>D1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(knot1 \leq t &lt; knot2)</td>
<td>D2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(knot2 \leq t &lt; knot3)</td>
<td>D3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(knot3 \leq t \leq n)</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
$int_1$ and $int_2$ are the predicted values of the observation at the change-point between phases A1 and B1 ($knot_1$), using the linear function in phases A1 and B1, respectively; $int_3$ and $int_4$ are the predicted values of the observation at the change-point between phases A2 and B2 ($knot_3$), using the linear function in phases A2 and B2, respectively. The predicted values of the observation at the change-point between phases B1 and A2 ($knot_2$) can be calculated using equations 16 and 17, using the linear function in phases B1 and A2, respectively. Figure 2 demonstrates the intercepts conceptually.

\begin{align*}
int_{2b} &= int_2 + slope_2 \times (knot_2 - knot_1) \\
int_{3b} &= int_3 - slope_3 \times (knot_3 - knot_2)
\end{align*}

(16)  

(17)

Figure 2. Intercepts for the linear model in the A-B-A-B design (equations 15, 16, and 17).

Posterior distributions of three knots (change-points estimation), four intercepts (level for each phase), three intercept changes, four slopes (trend for each phase), three slope changes, autocorrelation coefficient, and the standard deviation ($\sigma_\epsilon$, stability) are estimated with Bayesian methods. Model equations and prior distributions
can be found in the JAGS model code in Appendix C. Bayes factors (de Vries & Morey, 2013) can be evaluated for each of the two adjacent phases: A1B1; B1A2; and A2B2.

Examples

This section demonstrates the data analyses described above. We provide analyses of the data from Bernhardsson, Klintberg and Wendt (2011) for the baseline and the A-B design, and the data from Lambert, Cartledge, Heward and Lo (2006) for the A-B-A-B design.

Baseline Data Analysis

To illustrate the procedure of analyzing the baseline phase in SSR, we use the baseline data from Bernhardsson et al. (2011). The primary outcomes were intensity of shoulder pain, measured by a visual analogue scale (VAS) with scores ranging from 0 to 100 mm. We assume the data collection and analysis start from Session 1. Data analyzing should follow the process of data collection and therefore be conducted at every time point. In order to illustrate the process of conducting data analyses in the baseline phase, we demonstrate the data plotting and analyzing three times with 5, 8, and 11 sessions.

For the first 5 sessions in the baseline phase, we analyze the data with the R code in Appendix A. The plot is shown in Figure 3.
Figure 3. Plotting baseline data with the first 5 time points, Bernhardsson et al. (2011).

Visual analysis indicates fairly stable data pattern of a flat series with a flat trend. The measure of VAS stayed at a high level. Statistically, the stability and trend can be estimated by conducting the proposed baseline model. The trace and density plots of the parameter estimation are shown in Figures 4 and 5. The first section of Table 3 presents the output summary of the estimation for the parameters: rho (autocorrelation), int (intercept), slope, and sigma (standard deviation).
Figure 4. Trace plots for the monitored parameters (four chains) for baseline data with 5 time points.

Figure 5. Density plots for the monitored parameters (four chains) for baseline data with 5 time points.
The trace plots (Figure 4) and density plots (Figure 5) for four chains indicate convergence of parameter estimates. The estimates for the autocorrelation parameter (rho) show that the observations are indeed serially correlated. The mean of the estimates for the slope parameter is larger than 0, but the HDI includes 0, indicating that the trend is not statistically significantly different from zero.

For T = 8 and T = 11, with more baseline data collected, we plot the data again and repeat the same analysis procedure as for the first 4 sessions (Note: The number of observations should be changed to T = 8 and T = 11 in the R code). The total estimation took 5.7 seconds for 8 time points and 7.8 seconds for 11 time points. The output summaries are presented in the second and third sections in Table 3. The data plots, trace plots, and density plots for 8 and 11 sessions are in Figures 6 to 11.
Figure 6. Plotting baseline data with 8 time points, Bernhardsson et al. (2011).

Figure 7. Trace plots for the monitored parameters (four chains) for baseline data with 8 time points.
Figure 8. Density plots for the monitored parameters (four chains) for baseline data with 8 time points.

Figure 9. Plotting baseline data with 11 time points, Bernhardsson et al. (2011).
Figure 10. Trace plots for the monitored parameters (four chains) for baseline data with 11 time points.

Figure 11. Density plots for the monitored parameters (four chains) for baseline data with 11 time points.
The trace plots (Figures 7 and 10) and density plots (Figures 8 and 11) for four chains indicate convergence of parameter estimates. Comparing the parameter estimates for 5, 8, and 11 time points in Table 3, we notice that the standard deviations for the parameter estimates are lower for larger sample, indicating that more data points lead to better estimates. The output summaries (Table 3) and the posterior distributions indicate that the baseline data form a flat series with a flat trend (near zero slope). Comparing the estimates for the standard deviation parameter ($\sigma_e$, sigma) for 5, 8, and 11 time points, we notice that the estimates for the standard deviation parameters are stable through the baseline phase. This indicates stability of the baseline data.

Data Analysis of the A-B Design

Bernhardsson et al. (2011) evaluated the effect of eccentric strengthening exercises on pain intensity using an A-B design. The dependent variable was intensity of shoulder pain. It was measured by a visual analogue scale (VAS) with scores ranging from 0 to 100 mm. The intervention phase began on Session 12. In this paper, we demonstrate the data analysis at Session 34, with 11 baseline time points and 23 intervention phase time points. Figure 12 shows the pattern of the data.
The total estimation time was 1.5 minutes. The mode of the posterior distribution of the change-point parameter is 13, which is one time point behind the actual change-point. The proposed model detected the change-point of the A-B design with a delay effect of one time point. Table 4 presents the output summary of the estimation for the parameters: dint (level change), dslope (trend change), int1, int2, knot (change-point), rho (autocorrelation), sigma (standard deviation), slope1, and slope2. The trace plots and posterior density plots are shown in Figures 13 and 14, respectively.
### Table 4

**Bayesian Estimation Output Summary for the A-B Design Data Analysis**

<table>
<thead>
<tr>
<th>Param.</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>dslope</td>
<td>-0.640</td>
<td>0.479</td>
<td>-1.574</td>
<td>-0.646</td>
<td>0.323</td>
</tr>
<tr>
<td>intercept1</td>
<td>82.002</td>
<td>3.255</td>
<td>75.720</td>
<td>82.008</td>
<td>88.446</td>
</tr>
<tr>
<td>intercept2</td>
<td>58.298</td>
<td>2.522</td>
<td>53.368</td>
<td>58.332</td>
<td>63.336</td>
</tr>
<tr>
<td>knot</td>
<td>12.998</td>
<td>0.061</td>
<td>13.000</td>
<td>13.000</td>
<td>13.000</td>
</tr>
<tr>
<td>rho</td>
<td>0.103</td>
<td>0.087</td>
<td>0.000</td>
<td>0.082</td>
<td>0.273</td>
</tr>
<tr>
<td>sigma</td>
<td>5.470</td>
<td>0.717</td>
<td>4.172</td>
<td>5.398</td>
<td>6.908</td>
</tr>
<tr>
<td>slope1</td>
<td>-0.231</td>
<td>0.429</td>
<td>-1.098</td>
<td>-0.227</td>
<td>0.583</td>
</tr>
<tr>
<td>slope2</td>
<td>-0.871</td>
<td>0.204</td>
<td>-1.273</td>
<td>-0.873</td>
<td>-0.471</td>
</tr>
</tbody>
</table>

**Figure 13.** Trace plots for the monitored parameters (four chains) for the A-B design.
Figure 14. Density plots for the monitored parameters (four chains) for the A-B design.

The trace and density plots for four chains indicate convergence of parameter estimates. The estimates of the 95% HDI for the phase change-point parameter (knot) concentrate at the time point of 13. Based on the 4000 times estimation, 13 is the only time point that is most likely to be the point of phase change. The large amount of iteration and the high accuracy of the estimates provide additional statistical evidence for the effect of the intervention with 1 time point latency. The parameter estimates for the intercept change (dintercept) indicate that the level change is negative and statistically significantly different from 0. The average magnitude of the level change is 23.705. The negative sign indicates the decrease of the VAS. As soon as the intervention was introduced, the data showed large decrease in the dependent variable.
The slope change (dslope) is more likely to be negative, but the HDI includes 0 (-4.739, 3.480), indicating that the slope change is not statistically significantly different from 0.

Using the R codes in Appendix B for the Bayes factors by de Vries and Morey (2013), we test the hypothesis of level and trend changes. The Bayes factors for level change only, trend change only, and both level and trend changes are 0.170, 8.384, and 1.740, respectively. Therefore the ratios of the probability favoring the alternative hypotheses are 5.953, 0.119, and 0.575, for the three hypotheses. That is, compared to the null hypothesis of no level change, level change is about six times more likely. The slope change, however, is not supported. There is little evidence supporting the hypotheses of slope change only and both level and slope changes. Bayes factors agree with the estimation results.

The proposed model has detected the existence of change between the two phases. We have demonstrated that the data pattern change is from change of level, but not slope. Therefore, the statistical model has provided evidence for the intervention effect on the level change. Moreover, the proposed model has located the exact time point where the change occurred. The change took place at time point 13, which is one time point after the intervention was implemented. This indicates a delay effect of the intervention method on the target behavior.

Data Analysis for A-B-A-B Designs

Lambert et al. (2006) evaluated the effect of the use of response cards by school students using an A-B-A-B design. In this paper, we demonstrate the data analysis on the dependent variable of the number of disruptive behaviors. The phase changes took place at time points 9, 15, and 23. The data are presented in Figure 15.
When running the statistical model, the total estimation took 5.4 minutes. The modes of the posterior distributions of the change-point parameters are 10, 15, and 27, respectively. The model detected the first change-point at time point 10 while the actual change-point was at time point 9. This is because there is a missing data at time point 9. For the second change-point, the model detected the actual value accurately. The model detected the third change-point at the time point 27, instead of the actual change-point of 23. This shows a possible delay effect from the baseline phase A2 to the intervention phase B2. Table 5 presents the output summary of the estimation for the parameters: 3 knots (change-points), 4 intercepts, 3 intercept changes, rho (autocorrelation), sigma (standard deviation), 4 slopes, and 3 slope changes. The trace plots and posterior density plots are presented in Figures 16 and 17, respectively.

Figure 15. Plot for the A-B-A-B design data from Lambert et al. (2006).
Table 5

Bayesian Estimation Output Summary for A-B-A-B Design Data Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>dintercept1</td>
<td>-1.181</td>
<td>4.571</td>
<td>-8.884</td>
<td>-1.308</td>
<td>7.234</td>
</tr>
<tr>
<td>dintercept2</td>
<td>-2.215</td>
<td>30.434</td>
<td>-35.841</td>
<td>-2.652</td>
<td>29.539</td>
</tr>
<tr>
<td>dintercept3</td>
<td>-0.267</td>
<td>4.217</td>
<td>-7.651</td>
<td>-0.287</td>
<td>7.832</td>
</tr>
<tr>
<td>dslope1</td>
<td>0.471</td>
<td>3.753</td>
<td>-4.015</td>
<td>0.454</td>
<td>5.015</td>
</tr>
<tr>
<td>dslope2</td>
<td>0.101</td>
<td>3.990</td>
<td>-4.764</td>
<td>0.097</td>
<td>4.931</td>
</tr>
<tr>
<td>dslope3</td>
<td>-0.772</td>
<td>3.421</td>
<td>-2.768</td>
<td>-0.928</td>
<td>1.913</td>
</tr>
<tr>
<td>intercept1</td>
<td>5.581</td>
<td>2.898</td>
<td>0.073</td>
<td>6.202</td>
<td>9.558</td>
</tr>
<tr>
<td>intercept2</td>
<td>4.400</td>
<td>3.034</td>
<td>0.000</td>
<td>4.186</td>
<td>9.344</td>
</tr>
<tr>
<td>intercept3</td>
<td>6.123</td>
<td>2.759</td>
<td>0.921</td>
<td>6.617</td>
<td>10.000</td>
</tr>
<tr>
<td>intercept4</td>
<td>5.857</td>
<td>3.111</td>
<td>0.493</td>
<td>6.614</td>
<td>10.000</td>
</tr>
<tr>
<td>knot1</td>
<td>10.553</td>
<td>5.496</td>
<td>1.000</td>
<td>10.000</td>
<td>20.000</td>
</tr>
<tr>
<td>knot2</td>
<td>13.238</td>
<td>7.320</td>
<td>1.000</td>
<td>13.000</td>
<td>27.000</td>
</tr>
<tr>
<td>knot3</td>
<td>21.200</td>
<td>5.666</td>
<td>12.000</td>
<td>21.000</td>
<td>33.000</td>
</tr>
<tr>
<td>rho</td>
<td>0.207</td>
<td>0.153</td>
<td>0.000</td>
<td>0.177</td>
<td>0.503</td>
</tr>
<tr>
<td>sigma</td>
<td>2.429</td>
<td>0.469</td>
<td>1.605</td>
<td>2.365</td>
<td>3.367</td>
</tr>
<tr>
<td>slope1</td>
<td>-0.318</td>
<td>0.841</td>
<td>-1.561</td>
<td>-0.274</td>
<td>0.670</td>
</tr>
<tr>
<td>slope2</td>
<td>0.153</td>
<td>3.649</td>
<td>-4.366</td>
<td>0.206</td>
<td>4.261</td>
</tr>
<tr>
<td>slope3</td>
<td>0.254</td>
<td>1.611</td>
<td>-1.706</td>
<td>0.333</td>
<td>1.999</td>
</tr>
<tr>
<td>slope4</td>
<td>-0.518</td>
<td>3.002</td>
<td>-1.653</td>
<td>-0.570</td>
<td>0.870</td>
</tr>
</tbody>
</table>
Figure 16. Trace plots for the monitored parameters (four chains) for the A-B-A-B design.
Figure 17. Density plots for the monitored parameters (four chains) for the A-B-A-B design.
The trace and density plots for four chains indicate convergence of parameter estimates. The parameter estimates for the autocorrelation parameter indicate that the data are serially dependent as expected of SSR designs. The standard deviation estimate shows relatively low stability of the overall time series. The HDI of the slope1 estimate includes 0, indicating that the trend of the baseline phase A1 is not statistically significantly different from 0. The estimates for the rest of the parameters, however, have very wide HDI. Possible reasons are: (1) wide HDIs of the change-point parameters estimates and (2) the large SD of the time-series.

To sum up, the trend for each phase is not statistically significantly different from 0, indicating zero trend in each phase of the A-B-A-B design. The data pattern changes have been verified by the change-points estimates, however, the level and trend changes are not big enough to be detected. One possible reason is the large standard deviation of the time series. In other words, the stability is so low that the level and trend changes across phases became statistically less significant. This shows that stability of parameter estimates plays an important role in the data analysis.

Using the R codes for the Bayes factors (de Vries & Morey, 2013) in Appendix C, we calculate the ratio favoring the alternative hypothesis of the level and trend changes for each adjacent phases. The Bayes factors (BF) and the ratio favoring the alternative hypotheses (RA) are presented in Table 6. For the first phase change, there is no evidence for the hypothesis of trend change only. For the second phase change, there is no evidence for both of the hypotheses on level change only and the trend change only. The rest of the alternative hypotheses are supported by little evidence (RA of 1 or 2).
Table 6

Bayes Factors for Each of the Adjacent Phases in the A-B-A-B Design

<table>
<thead>
<tr>
<th>Phases</th>
<th>Level change only</th>
<th>Trend change only</th>
<th>Both level and trend changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BF(^a)</td>
<td>RA(^b)</td>
<td>BF</td>
</tr>
<tr>
<td>A1 B1</td>
<td>0.849</td>
<td>1.178</td>
<td>3.307</td>
</tr>
<tr>
<td>B1 A2</td>
<td>1.482</td>
<td>0.675</td>
<td>3.175</td>
</tr>
<tr>
<td>A2 B2</td>
<td>0.490</td>
<td>2.040</td>
<td>0.485</td>
</tr>
</tbody>
</table>

Note. \(^a\) BF: Bayes Factor (de Vries & Morey, 2013). \(^b\) RA: Ratio favoring the alternative hypothesis (RA = 1/BF).

Discussion

Causal effects have been the interest of social scientists along the history of social sciences (Bollen & Pearl, 2013). The ideal research condition to prove causality is an isolated experimental environment where all the spurious variables are controlled. However, isolation is not always possible to accomplish in applied research, just like how difficult it is to control all threats to internal validity. The proposed statistical model provides additional evidence of causal relationship. We estimate the change-point as an unknown parameter. For the change-point parameter, we use a less informative prior distribution with equal probability for all the time points. Based on the data pattern of the time-series, this approach provides evidence for (1) the existence of changes between phases, (2) the time point when the changes take place and, (3) pattern of the changes on level and trend. When the change-point parameter is estimated with accuracy, the switch between the baseline and intervention phases is detected based on the pattern of the observed data. This allows the data to speak for itself, confirms presence of intervention effect by detecting the time of change. When the estimated change-point is behind its actual value, it may indicate a delay effect of the intervention method. On the
other hand, change-point parameter with wide HDI indicates little or no evidence of the intervention effect.

Additionally, pattern of change is described by the measurements of level and trend changes. We have presented the procedure for analyzing posterior distributions of level and trend change parameters. Although previous studies have used Bayesian methods in the analysis of level and trend changes, we focus on analyzing the quality of the parameter estimates by studying the posterior distributions. Analyzing posterior distributions provides rich information about level and trend changes and draws a clearer picture of the changing patterns. We recommend that inference be based on the comparison between the highest density intervals (HDI) of the posterior distributions and the regions of practical equivalence (ROPE) of the level and trend change parameters.

This article describes a set of linear modeling approaches that (a) estimate the change-points as unknown parameters to evaluate the effect of the intervention, (b) address the problem of serial dependency and, (c) present a Bayesian posterior distribution analysis procedure for three types of single subject designs (SSDs). Current statistical methods for analysis of SSDs have considered level and trend changes at the point of phase change. To the best of our knowledge, this is the first time that the change-point is treated as an unknown parameter in the statistical model. Change-point estimation and level and trend changes provide complementary information and, when used in conjunction, can provide a better understanding of the intervention effect.

In defining and describing the change-point evaluation approach for SSDs, we have provided analysis methods for the designs with two and four phases. For the A-B
design, the change-point estimate can be used directly to draw inferences about the intervention effect. For the A-B-A-B design, we have defined an approach that can estimate all three change-points.

The approaches that we have developed in this study are based on the linear regression model. We address the problem of serial dependency by taking autocorrelation into account. Not considering autocorrelations may lead to incorrect results and misinterpretation of the intervention effect. The fact that the proposed models consider autocorrelation meets the demand of time-series type of data and provides a more accurate evaluation of the intervention effect.

We calculated Bayes factors for the data analysis examples of the A-B and A-B-A-B designs. We also calculated the ratio favoring the alternative hypothesis (RA), that is, the inverse of Bayes factor. Larger RA shows more evidence of the level and trend changes. The proposed analysis approach is superior. Results from the BF and RA have confirmed the information provided by the proposed approach, but do not provide information about level and trend differences in sufficient detail. In addition to the statistical significance of the data pattern change, the approach described in the present study also provides information about the magnitude of the level and trend differences between adjacent phases. Just as the t test provides information about statistical significance of differences between groups, the proposed Bayesian data analysis approach provides evidence for statistical significance of changes between phases.

In single subject designs (SSDs), threats to internal validity may be controlled by repeated measurements taken during the baseline phase (Engel & Schutt, 2009). Problems of maturation, statistical regression, measurement effects (a.k.a. testing
effects), and instrumentation may be controlled by achieving a stable pattern of the
dependent variable. When the baseline phase shows a non-zero trend, especially in the
desired direction, it is more difficult to demonstrate the effectiveness of the intervention.
The change-point estimation methods have provided an approach that loosens the
requirement of flat baseline and provides additional evidence for causality.

Like any study, the present study has some limitations. Several effect sizes are
available in SSR, however, this study only considered one effect size, the Bayes factor.
The Bayes factors calculated by the R package “BayesSingleSub” can only be used for
designs with two phases. For the A-B-A-B design, we applied the Bayes factors three
times for each two adjacent phases. The present study does not provide a single effect
size measure for the entire four phases of the A-B-A-B design. This needs further
investigation.

In the second manuscript of this dissertation, we continue exploring the data
analysis approaches for three other SSDs: multiple baseline design, alternating
treatments design, and changing criterion design. The variety of different types of
designs and the detailed description of the programing codes provide single subject
researchers a set of tools that can be applied in the statistical analysis of SSR.
References


Appendix A

R and JAGS Codes for the Baseline Data Analysis
JAGS Code (Baseline.txt):

model
{
  Yhat[1] <- slope + int
  Y[1] ~ dnorm(Yhat[1], tau.e)
  tau.e <- 1/pow(sigma.e,2)

  for (t in 2:T){
    Yhat[t] <- (1-rho)*int + (t-rho*(t-1))*slope + rho*Y[t-1]
    Y[t] ~ dnorm(Yhat[t], tau.delta)
  }

  sigma <- 1/sqrt(tau.delta)
  sigma.e <- sigma/sqrt(1-pow(rho,2))

  int ~ dunif(ymin-slope,ymax)
  slope ~ dnorm(0,precs[1])
  precs[1] ~ dgamma(1,1)
  rho ~ dbeta(1,5)
  tau.delta ~ dgamma(1,1)

  #data# T, Y, ymin, ymax
}

R Code (Baseline.R):

install.packages('rjags')                           #Install and load R packages
install.packages('runjags')
require(rjags)
require(runjags)
setwd('E:/Dissertation/Data Analysis/Baseline and AB')    #Set working directory
data <- as.matrix(read.csv('Bernhardsson et al 2011.csv', header=T))   #Load data file
T = 5   #specity the number of observations in Baseline phase - 5, 8, or 11
T = 8
T = 11
Y = as.numeric(data[seq(1,T),2])                   #specify baseline data
ymin <- min(Y, na.rm = T) #min and max of Y, missing data removed
ymax <- max(Y, na.rm = T)
plot(Y, type="o", bty="i", pch=16,                  #Plot baseline data
     xlim=c(0,1.5*T), ylim=c(0,ymax), xlab="Sessions", ylab="visual analogue scale")
abline(abline(v = T + 0.5, lty = 1))
mtext("Baseline", side=3, line=0, outer=F, adj=0, cex=1)
#Run the analysis, monitoring slope/intercept/standard deviation/autocorrelation
runjags.object <- autorun.jags("Baseline.txt", monitor = c("slope", "int",
                                       "sigma", "rho"),
                                   n.chains=4, startburnin = 25000, startsample = 4000,
                                   thin.sample = FALSE, thin = 1)
plot(runjags.object, type="trace", layout=c(2,2))
plot(runjags.object, type="density", layout=c(2,2))
runjags.object                                      #Output Summary

Data file (Bernhardsson et al 2011.csv):

<table>
<thead>
<tr>
<th>Sessions</th>
<th>VAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>82</td>
</tr>
<tr>
<td>4</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>91</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>84</td>
</tr>
<tr>
<td>8</td>
<td>86</td>
</tr>
<tr>
<td>9</td>
<td>79</td>
</tr>
<tr>
<td>10</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>81</td>
</tr>
<tr>
<td>12</td>
<td>81</td>
</tr>
<tr>
<td>13</td>
<td>55</td>
</tr>
<tr>
<td>14</td>
<td>59</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td>17</td>
<td>58</td>
</tr>
<tr>
<td>18</td>
<td>55</td>
</tr>
<tr>
<td>19</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>51</td>
</tr>
<tr>
<td>21</td>
<td>69</td>
</tr>
<tr>
<td>22</td>
<td>51</td>
</tr>
<tr>
<td>23</td>
<td>43</td>
</tr>
<tr>
<td>24</td>
<td>49</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>26</td>
<td>40</td>
</tr>
<tr>
<td>27</td>
<td>39</td>
</tr>
<tr>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>29</td>
<td>40</td>
</tr>
<tr>
<td>30</td>
<td>41</td>
</tr>
<tr>
<td>31</td>
<td>50</td>
</tr>
<tr>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td>34</td>
<td>35</td>
</tr>
</tbody>
</table>
Appendix B

R and JAGS Codes for the A-B Design Data Analysis
**JAGS Code (AB.txt):**

```jags
data{
  T, Y, ymax
}

model{
  Yhat[1] <- slope1 + (int1 - slope1 * knot)
  Y[1] ~ dnorm(Yhat[1], tau.e)
  p[1] <- 1/T
  D[1] <- 0

  for (t in 2:T){
    p[t] <- 1/T
    D[t] <- step(t - knot)
    Yhat[t] <- ((1 - D[t]) - rho * (1 - D[t-1])) * (int1 - slope1 * knot) + ((1 - D[t]) * t - rho * (1 - D[t-1]) * (t-1)) * slope1
    + (D[t] - rho * D[t-1]) * (int2 - slope2 * knot) + (D[t] * t - rho * D[t-1] * (t-1)) * slope2 + rho * Y[t-1]
    Y[t] ~ dnorm(Yhat[t], tau.delta)
  }

  tau.e <- 1/pow(sigma.e,2)
  sigma <- 1/sqrt(tau.delta)
  sigma.e <- sigma/sqrt(1 - pow(rho,2))

  int1 ~ dunif(0,ymax)
  int2 ~ dunif(0,ymax)
  dint <- int2 - int1
  slope1 ~ dnorm(0,precs[1])
  slope2 ~ dnorm(0,precs[2])
  dslope <- slope2 - slope1
  precs[1] ~ dgamma(1,1)
  precs[2] ~ dgamma(1,1)
  knot ~ dcat(p[])
  tau.delta ~ dgamma(1,1)
  rho ~ dbeta(1,5)
}
```

**R Code (AB.R):**

```r
install.packages('rjags')               #Install and load R packages
install.packages('runjags')
install.packages('BayesSingleSub')
install.packages('prettyR')
require(rjags)
require(runjags)
require(BayesSingleSub)
```

45
require(prettyR)
setwd("/Volumes/LEXAR/Dissertation/Data Analysis/Baseline and AB")    #Set working directory
data <- as.matrix(read.csv('Bernhardsson et al 2011.csv', header=T)) #Load data file
T <- nrow(data)
Y <- as.numeric(data[,2])  #maxima of the input values, missing values removed
ymax <- max(Y, na.rm = T) #maxima of the input values, missing values removed

plot(Y, type="o", bty="l", pch=16, ylim=c(0,ymax),   #Plot the data  
cex.lab=0.8, xlab="Weeks", ylab="visual analogue scale")
abline(v = 11.5, lty = 1)
mtext("Baseline", side=3, line=0, outer=F, adj=0, cex=0.8)
mtext("Intervention", side=3, line=0, outer=F, adj=0.75, cex=0.8)

#Run the analysis, monitor slopes/intercepts/standard deviation/autocorrelation/knot  
runjags.object <- autorun.jags("AB.txt", monitor = c("slope1", "slope2", "dslope",  
"int1", "int2", "dint", 
"knot", "rho", "sigma"),
  n.chains=4, startburnin = 25000, startsample = 4000,  
  thin.sample = FALSE, thin = 1)

knot.est <- as.mcmc.list(runjags.object,vars="knot")
knot.est <- as.matrix(knot.est)
knot.mode <- Mode(knot.est)
runjags.object   #Output Summerry

plot(runjags.object, type="trace", layout=c(3,3))   #trace plot  
plot(runjags.object, type="density", layout=c(3,3)) #density plot
plot(runjags.object, type="all", layout=c(7,2))    #plots organized by parameter

#Calculate (de Vries and Morey's) Bayes factors
n1 = 11   #number of observations in Baseline phase
n2 = T-n1
ypre = Y[1:n1]
ypress = Y[n1 + 1:n2]

#(mean difference - only a sudden stable level shift)
logBAR = ttest.Gibbs.AR(ypre, ypost, iterations = 10000, return.chains = FALSE,  
r.scale = 1, betaTheta = 5, sdMet = 0.3)

exp(logBAR)  #Bayes factor (Bar, ratio favoring the null)
1/exp(logBAR) #ratio favoring the alternative

#to examine the posterior distribution for any parameter
output.bar = ttest.Gibbs.AR(ypre, ypost, iterations = 10000, return.chains = TRUE, 
   r.scale = 1, betaTheta = 5, sdMet = 0.3, 
   areaNull = c(-0.2, 0.2))
exp(output.bar$logbf) #Bayes factor (Bar)
1/exp(output.bar$logbf)
#summary of: mu0, grand mean; delta, effect size; sig2, variance; rho, autocorrelation
summary(output.bar$chains)
plot(output.bar$chains) #need enough space to show plots

#calculating Bayes factor for mean & trend difference
#(Bi+t, Bint, Btrend - assuming trend exists in the data pattern)
logBTRENDS = trendtest.Gibbs.AR(ypre, ypost, iterations = 10000, 
   return.chains = FALSE, r.scaleInt = 1, r.scaleSlp = 1, 
   betaTheta = 5, sdMet = 0.3)
exp(logBTRENDS) #Bayes factors (Bi+t, Bint, Btrend, ratio favoring the null)
1/exp(logBTRENDS) #ratio favoring the alternative
#to examine the posterior distribution for any parameter
output.trend = trendtest.Gibbs.AR(ypre, ypost, iterations = 10000, 
   return.chains = TRUE, r.scaleInt = 1, r.scaleSlp = 1, 
   betaTheta = 5, sdMet = 0.3, intArea = c(-0.2, 0.2), 
   slpArea = c(-0.1, 0.1))
exp(output.trend$logbfArea) #Bint, Btrend
1/exp(output.trend$logbfArea) #ratio favoring the alternative
#summary of: beta0, general trend; sig*beta1, standardized trend difference
summary(output.trend$chains)
plot(output.trend$chains)
Appendix C

R and JAGS Codes for the A-B-A-B Design Data Analysis
JAGS Code (ABAB.txt):

define "model"
{
    Yhat[1] <- slope1 + (int1 - slope1 * knot1)
    Y[1] ~ dnorm(Yhat[1], tau.e)
    D1[1] <- 0
    D2[1] <- 0
    D3[1] <- 0

    for (t in 2:T){
        D1[t] <- step(t - knot1)
        D2[t] <- step(t - knot2)
        D3[t] <- step(t - knot3)
        Yhat[t] <- ((1 - D1[t]) - rho * (1 - D1[t-1])) * (int1 - slope1 * knot1)
            + ((1 - D1[t]) * t - rho * (1 - D1[t-1]) * (t-1)) * slope1
            + (D1[t] * (1 - D2[t]) - rho * D1[t-1] * (1 - D2[t-1])) * (int2 - slope2 * knot1)
            + (D1[t] * (1 - D2[t]) * t - rho * D1[t-1] * (1 - D2[t-1]) * (t-1)) * slope2
            + (D1[t] * D2[t] * (1 - D3[t]) - rho * D1[t-1] * D2[t-1] * (1 - D3[t-1])) * (int3 - slope3 * knot3)
            + (D1[t] * D2[t] * (1 - D3[t]) * t - rho * D1[t-1] * D2[t-1] * (1 - D3[t-1]) * (t-1)) * slope3
            + (D1[t] * D2[t] * D3[t] - rho * D1[t-1] * D2[t-1] * D3[t-1]) * (int4 - slope4 * knot3)
            + (D1[t] * D2[t] * D3[t] * t - rho * D1[t-1] * D2[t-1] * D3[t-1] * (t-1)) * slope4
            + rho * Y[t-1]
        Y[t] ~ dnorm(Yhat[t], tau.delta)
    }

    tau.e <- 1/pow(sigma.e,2)
    sigma.e <- sigma/sqrt(1 - pow(rho,2))
    sigma <- 1/sqrt(tau.delta)
    int2b <- int2 + slope2 * (knot2 - knot1)
    int3b <- int3 - slope3 * (knot3 - knot2)
    dint1 <- int2 - int1
    dint2 <- int3b - dint1
    dint3 <- int4 - dint3
    dslope1 <- slope2 - slope1
    dslope2 <- slope3 - slope2
    dslope3 <- slope4 - slope3

    int1 ~ dunif(0,ymax)
    int2 ~ dunif(0,ymax)
    int3 ~ dunif(0,ymax)
    int4 ~ dunif(0,ymax)
    slope1 ~ dnorm(0,precs[1])
    slope2 ~ dnorm(mu,precs[2])
}
mu ~ dnorm(0, sprec)
sprec ~ dgamma(1, 1)
slope3 ~ dnorm(0, precs[3])
slope4 ~ dnorm(0, precs[4])
precs[1] ~ dgamma(1, 1)
precs[2] ~ dgamma(1, 1)
precs[3] ~ dgamma(1, 1)
precs[4] ~ dgamma(1, 1)

for (t in 1:T){
  p[t] <- 1/T
}
knot1 ~ dcat(p[])
knot2 ~ dcat(p[])
knot3 ~ dcat(p[])
tau.delta ~ dgamma(1, 1)
rho ~ dbeta(1, 5)
#data# T, Y, ymax
}

R Code (ABAB.R):

install.packages('rjags')               #Install and load R packages
install.packages('runjags')
install.packages('prettyR')
install.packages('BayesSingleSub')
require(rjags)
require(runjags)
require(prettyR)
require(BayesSingleSub)
setwd('/Volumes/LEXAR/Dissertation/Data Analysis/ABAB')          #Set working directory
data <- as.matrix(read.csv('Lambert et al 2006C1.csv', header = T)) #Load class 1 data
T <- nrow(data)
Y <- as.numeric(data[,4]) #student 3
ymax <- max(Y, na.rm = T)
plot(Y, type="o", bty="l", pch=16, ylim=c(0,ymax), cex.lab=0.8,    #Plot the data
     xlab="Sessions", ylab="#Disruptive Behavior")
abline(v = 8.5, lty = 3)
abline(v = 14.5, lty = 3)
abline(v = 22.5, lty = 3)

#Run the analysis, monitoring interested parameters
runjags.object <- autorun.jags("ABAB.txt", monitor = c("slope1", "slope2", "slope3",
          "slope4",
          "int1", "int2", "int3", "int4"),
"dint1", "dint2", "dint3", "dslope1", "dslope2", "dslope3",
"knot1", "knot2", "knot3", "rho", "sigma"),
n.chains=4, startburnin = 25000, startsample = 4000, thin.sample =
FALSE, thin = 1)
#mode for the three change-points
knot1.est <- as.mcmc.list(runjags.object, vars="knot1")
knot1.est <- as.matrix(knot1.est)
knot1.mode <- Mode(knot1.est)
knot2.est <- as.mcmc.list(runjags.object, vars="knot2")
knot2.est <- as.matrix(knot2.est)
knot2.mode <- Mode(knot2.est)
knot3.est <- as.mcmc.list(runjags.object, vars="knot3")
knot3.est <- as.matrix(knot3.est)
knot3.mode <- Mode(knot3.est)
knot1.mode
knot2.mode
knot3.mode
runjags.object #output summary
plot(runjags.object, type="trace", layout=c(7,3))
plot(runjags.object, type="density", layout=c(7,3))

#Calculate (de Vries and Morey's) Bayes factors
n1 = 8  #Specify the number of observations in each phase
n2 = 6
n3 = 8
n4 = 8
ypre = Y[1:n1]
ypost = Y[n1 + 1:n2]
ypost1 = Y[n1 + n2 + 1:n3]
ypost2 = Y[n1 + n2 + n3 + 1:n4]
# Bayes factor for mean &/ trend difference (Bi+t, Bint, Btrend - assuming trend exists in
the data pattern)
#between phases A1 and B1
logBTRENDS = trendtest.Gibbs.AR(ypre, ypost, iterations = 10000, return.chains =
FALSE, r.scaleInt = 1, r.scaleSlp = 1, betaTheta = 5, sdMet = 0.3)
exp(logBTRENDS) #Bayes factors (Bi+t, Bint, Btrend, ratio favoring the null)
1/exp(logBTRENDS) #ratio favoring the alternative

#between phases B1 and A2
logBTRENDS1 = trendtest.Gibbs.AR(ypost, ypost1, iterations = 10000, return.chains =
FALSE, r.scaleInt = 1, r.scaleSlp = 1, betaTheta = 5, sdMet = 0.3)
exp(logBTRENDS1) #Bayes factors (Bi+t, Bint, Btrend, ratio favoring the null)
1/exp(logBTRENDS1) #ratio favoring the alternative

#between phases A2 and B2
logBTRENDS2 = trendtest.Gibbs.AR(ypost1, ypost2, iterations = 10000, return.chains = FALSE, r.scaleInt = 1, r.scaleSlp = 1, betaTheta = 5, sdMet = 0.3)

exp(logBTRENDS2) #Bayes factors (Bi+t, Bint, Btrend, ratio favoring the null)
1/exp(logBTRENDS2) #ratio favoring the alternative

Data file (Lambert et al 2006C1.csv):

<table>
<thead>
<tr>
<th>Sessions</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>26</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
BAYESIAN ANALYSIS OF MULTIPLE BASELINE DESIGN, ALTERNATING TREATMENTS DESIGN, AND CHANGING CRITERION DESIGN

Single subject research (SSR) data contains successive observations and consequently, is time series type of data (Barlow, Nock, & Hersen, 2009). In the first manuscript of this dissertation, we have proposed a Bayesian estimation approach based on time series data analysis. Serial dependency has been taken into consideration in the linear model. Additionally, a new change-point estimation approach has been introduced in the SSR data analysis. Data analysis procedures for the baseline phase, A-B design and A-B-A-B design have been the focus. In the current paper, we continue exploring Bayesian data analysis methods for three other SSR designs.

Bayesian estimation methods have already been applied in SSR data analyses. Zucker, Schmid, McIntosh, D'Agostino, Selker, and Lau (1997) applied a hierarchical Bayesian random effect model to combine single patient studies under the randomized controlled multi-crossover design. They applied the structured design for the assessment of medical treatment on chronic diseases. Swaminathan, Roger and Horner (2014) applied Bayesian method in a multilevel model, while taking into account the autocorrelation of error terms. de Vries and Morey (2013) introduced a set of Bayes factor tests for single subject data with two phases. The two latter studies took into consideration serial dependency.

However, de Vries and Morey's Bayes factors can only be applied on the two phases design or the adjacent phases of other designs. Swaminathan's approach is only applicable for the A-B and A-B-A-B designs with multiple subjects. None of these
studies have extended Bayesian data analyses to other commonly used designs in SSR. Multiple baseline design is the most popular SSR design (Shadish & Sullivan, 2011). Researchers seeking to compare different treatments need to use alternating treatments design (ATD, Wolery, Gast, & Hammond, 2010). Changing criterion design can be applied to instructional behaviors, monitoring the monotonic development of the target behavior (Gast & Ledford, 2010b). Both multiple baseline and changing criterion designs are able to verify replication effect. There is a need for data analysis procedures for these commonly used SSR designs.

Using Bayesian estimation of linear models, this paper demonstrates methods for three single subject designs (SSD): the multiple baseline design, ATD, and changing criterion design. Insight gained from such a study would be helpful for single subject researchers seeking to explain the intervention effect when using these three designs. The current study seeks to answer the following research question:

- How can Bayesian methods be used to analyze SSR data obtained using the multiple baseline design, ATD, and changing criterion designs?

The models developed in this study are confirmatory models. They are not used for prediction purpose. We demonstrate the utility of these models by applying them to the data analysis of empirical SSR studies from published works.

Methods

The current study demonstrates Bayesian data analysis methods for three SSR designs. The parameters will be estimated using Markov chain Monte Carlo (MCMC) methods through the software programs JAGS 3.4.0 (Plummer, 2003) and R 3.1.0 (R Core Team, 2014).
Multiple Baseline Design

Multiple baseline design is the most commonly used single subject research design (Shadish & Sullivan, 2011). It collects data across three or more tiers, which can be applied to different behaviors, conditions, or participants (Gast & Ledford, 2010a). In each tier, the model is similar to the A-B design, with two phases and one change-point. Assume the number of tiers in the multiple baseline design is \( i (i \geq 3) \). The model equations for the multiple baseline design are:

\[
\hat{y}_{i,1} = (int_{1i} - \beta_1 i \times knot_i) + \beta_1 i, \quad (1)
\]

\[
y_{i,1} \sim \text{dnorm}(\hat{y}_{i,1}, \tau_{ei}), \quad \tau_{ei} = \frac{1}{\sigma_{ei}^2}
\]

\[
\hat{y}_{i,t} = (int_{1i} - \beta_1 i \times knot_i)\left[(1 - D_{i,t}) - \rho_i (1 - D_{i,t-1})\right] + \beta_1 i\left[(1 - D_{i,t}) - \rho_i (1 - D_{i,t-1})(t - 1)\right] + (int_{2i} - \beta_1 i \times knot_i)(D_{it} - \rho_i D_{i,t-1})
\]

\[
+ \beta_2 i[D_{i,t}t - \rho_i D_{i,t-1}(t - 1)] + \rho_i y_{i,t-1}, \quad (t \geq 2)
\]

\[
y_{i,t} \sim \text{dnorm}(\hat{y}_{i,t}, \tau_{di}), \quad \tau_{di} = \frac{1}{\sigma_{di}^2}, \quad \sigma_{ei} = \frac{\sigma_{ei}}{\sqrt{1 - \rho_i^2}}
\]

In equation 2, \( D_{i,t} \) is the dummy variable. For each tier \( i \), \( D_{i,t} = 0 \) in the baseline phase (when \( t < knot_i \)); and \( D_{i,t} = 1 \) in the intervention phase (when \( t \geq knot_i \)).

For each tier \( i \), the monitored parameters include the standard deviation (\( \sigma_{i,x} \)), change-point (\( knot_i \)), autocorrelation (\( \rho_i \)), two slopes (\( \beta_1 i \) and \( \beta_2 i \)), slope change (\( \beta_2 i - \beta_1 i \)), two intercepts (\( int_{1i} \) and \( int_{2i} \); in order to compare the level changes at the change-points, the intercepts are set up as the predicted values at the point of \( knot \)), and intercept change (\( int_{2i} - int_{1i} \)). Prior distributions can be found in the JAGS model code in the Appendix. Corresponding codes for R and JAGS are in Appendix D. For the
baseline phase of each tier, it is necessary to apply the baseline data analysis as demonstrated in the first manuscript of this dissertation.

Alternating Treatments Design (ATD)

ATD is used to compare the relative effectiveness of two or more interventions. In ATD, the baseline phase is recommended but not mandatory, so does the best alone phase (phase III), where only the superior intervention method is retained (Wolery et al., 2010). We assume that $i (i \geq 2)$ intervention methods are applied. Different interventions share the same baseline data. In the method section, we describe the statistical models for ATD with two and three phases. The data analysis procedures for the two and three phase designs are similar. Therefore, we only demonstrate an example of the two-phase design in the next section.

The basic model for a two-phase design is presented in equation 3.

$$y_{i,t} = (1 - D_t)[(int_{1,i} - \beta_{1,i} \times knot_1) + \beta_{1,i}t]$$

$$+ D_t[(int_{2,i} - \beta_{2,i} \times knot_1) + \beta_{2,i}t] + e_{i,t}$$

$D_t$ is a dummy variable. In the baseline phase, $D_t = 0$ (when $t < knot$); and in the intervention phases, $D_t = 1$ (when $t \geq knot$).

The basic model for a three-phase design is presented in equation 4.

$$y_{i,t} = (1 - D_{1t})[(int_{1,i} - \beta_{1,i} \times knot_1) + \beta_{1,i}t]$$

$$+ D_{1t}(1 - D_{2t})[(int_{2,i} - \beta_{2,i} \times knot_1) + \beta_{2,i}t]$$

$$+ D_{1t}D_{2t}[(int_{3,i} - \beta_{3,i} \times knot_2) + \beta_{3,i}t] + e_{i,t}$$

$D_{1t}$ and $D_{2t}$ are dummy variables. Table 1 shows the value for each dummy variable at each phase. In phase I, $D_{1t} = 0$ (when $t < knot1$); and in phases II and III, $D_{1t} = 1$.
(when \( t \geq \text{knot1} \)). In phases I and II, \( D_2t = 0 \) (when \( t < \text{knot2} \)); and in phases III, \( D_2t = 1 \) (when \( t \geq \text{knot2} \)).

Table 1

**Values of Dummy Variables for Each Phase for the ATD Design**

<table>
<thead>
<tr>
<th></th>
<th>Baseline (phase I)</th>
<th>Intervention (phase II)</th>
<th>Best alone (phase III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_1 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( D_2 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Additionally \( i_{t,1} \) and \( i_{t,2} \) are the predicted values of the observation at the change-point between phases I and II (\( \text{knot1} \)), using the linear function in phases I and II, respectively; \( i_{t,3} \) is the predicted value of the observation at the change-point between phases II and III (\( \text{knot2} \)). The predicted values of the superior intervention at the change-point between phases II and III (\( \text{knot2} \)) can be calculated using equation 5.

Figure 1 demonstrates the intercepts conceptually.

\[
\text{int2}_{bi} = \text{int2}_i + \text{slope2}_i(\text{knot2} - \text{knot1})
\] (5)

Figure 1. Intercepts for the linear model in the ATD design (equations 20 and 21).
In the data file, the multiple intervention methods share the same baseline data. For the sessions-across-days design, where different interventions are alternately presented during phase two, the data should still be presented sequentially.

In ATD, stability is not a requirement (Wolery et al., 2010). Therefore the stability examination is not necessary in the ATD data analysis. With the focus of comparing the effects of different interventions, we can compare the intercepts and slopes in the second phase. For example, we compare \( \text{int}_{1,2} \) with \( \text{int}_{2,2} \) in Figure 1 to evaluate the level difference of treatments 1 and 2. The change-point between phases II and III can be used as an indicator of the persistence of the effect. When the change-point is not detected by the statistical model, the difference between the two phases is not readily apparent and the persistence of the effect is confirmed. Comparing the intercepts and slopes between the two phases will quantify the persistence of the effect of the superior intervention method. For example, in Figure 1, we compare \( \text{int}_{1,2b} \) with \( \text{int}_{1,3} \) for the persistence of the first treatment’s effect on the intercept.

Model equations for the two and three phase designs and corresponding prior distributions can be found in the JAGS model codes. R and JAGS codes for ATD are in Appendix E.

Changing Criterion Design

Changing criterion design is used for instructional behaviors. Following a baseline which shows stability and appropriate trend, a series of criterion changes were applied to the target behavior. Therefore, the behavior would show a sequence of corresponding changes (increases or decreases, depending on the direction of the criterion changes). The systematic stepwise criterion changes pushed the behavior
toward the same direction as the change of the criterion. On the other hand, each phase (step) plays the role of baseline for the next phase (step).

In the changing criterion design, the criterion plays the role of intervention. In each phase, obtaining stability is very important. The method introduced in the baseline data analysis can be repeatedly applied in each step (phase) of the changing criterion design. The purpose of the changing criterion design is to change the behavior monotonically, either in the increasing or decreasing direction. Therefore the overall model can be estimated with a linear regression model. Using all the phases (steps) through the entire time series, the linear model in equations 6 and 7 can be used to test the overall effect of the changing criterion technique.

\[ \hat{y}_1 = \text{int} + \beta_1, \]  

\[ \hat{y}_t = \text{int}(1 - \rho) + \beta_1[t - \rho(t - 1)] + \rho y_{t-1}, \quad (t \geq 2) \]  

In this paper we focus on the function of the linear models. However, if the data does not show a linear function according to the scatter plot of the time series, equations 6 and 7 should be replaced by appropriate polynomial functions.

For certain designs when researchers would like to test the replication effect by manipulating the criterion back to a previous level, the statistical model is analogous to the A-B design. Figure 2 uses the teaching sample from Gast and Ledford (2010b) to demonstrate the linear structure of this type of changing criterion design.

The JAGS model code for the changing criterion design is identical to the linear model presented in the baseline analysis or the two phases model in the first manuscript. Corresponding R and JAGS codes for the changing criterion design data analysis and plotting are in Appendix F.
In the next section, the R and JAGS program codes for the multiple baseline design, ATD, and changing criterion design are tested using empirical SSR data from published works of applied single subject research.

Examples

To illustrate the procedures described above, we provide analyses of the data used by Rodriguez and Anderson (2014) for the multiple baseline design, data by DeVeney, Cress, and Reid (2014) for the ATD, and data by Cameron, Shapiro and Ainsleigh (2005) for the changing criterion design.

Data Analysis for Multiple Baseline Designs

In the data used by Rodriguez and Anderson (2014), the phase changes for the five participant groups occurred at time points 7, 11, 16, 20 and 20. The data are presented in Figure 3.
Figure 3. Plot for the multiple baseline design data from Rodriguez and Anderson (2014).

The estimation took 57.9 minutes. The modes of the posterior distributions of the change-point parameters are 7, 12, 16, 3, and 9, respectively. The change-points for the first and third tiers have been detected accurately, indicating strong evidences for the intervention effects on Deborah and Barbara’s groups. The change-point for the second
tier has been detected one time point after the intervention was implemented. This indicates that Amy’s group reacted to the intervention with one time point delay. The change-points for the last two tiers failed to stand out correctly. This indicates that the intervention effects on Natasha and Candice’s groups were not statistically significant.

Table 2 presents the output summary of the estimation for the parameters (for each tier [i]): 2 intercepts (int1 and int2), intercept change (dint), change-point (knot), autocorrelation (rho), standard deviation (sigma), 2 slopes, and slope change (dslope).

The trace and density plots are shown in Figures 4 and 5, respectively.

Table 2

Bayesian Estimation Output Summary for the Multiple Baseline Design Data Analysis (Tiers [1], [2], [3], [4], and [5])

<table>
<thead>
<tr>
<th>Param.</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>dslope[1]</td>
<td>0.622</td>
<td>1.465</td>
<td>-0.050</td>
<td>0.388</td>
<td>3.892</td>
</tr>
<tr>
<td>dslope[2]</td>
<td>-1.387</td>
<td>0.799</td>
<td>-2.972</td>
<td>-1.353</td>
<td>0.081</td>
</tr>
<tr>
<td>dslope[3]</td>
<td>-0.992</td>
<td>0.744</td>
<td>-2.384</td>
<td>-0.988</td>
<td>0.538</td>
</tr>
<tr>
<td>dslope[4]</td>
<td>-0.815</td>
<td>2.003</td>
<td>-4.067</td>
<td>-0.723</td>
<td>3.044</td>
</tr>
<tr>
<td>dslope[5]</td>
<td>-0.992</td>
<td>2.023</td>
<td>-5.036</td>
<td>-0.880</td>
<td>2.395</td>
</tr>
<tr>
<td>intercept[1][1]</td>
<td>44.112</td>
<td>9.389</td>
<td>23.870</td>
<td>45.851</td>
<td>59.906</td>
</tr>
<tr>
<td>intercept[2][1]</td>
<td>37.857</td>
<td>4.782</td>
<td>28.739</td>
<td>37.841</td>
<td>47.254</td>
</tr>
<tr>
<td>intercept[3][1]</td>
<td>58.035</td>
<td>5.580</td>
<td>47.260</td>
<td>58.186</td>
<td>68.899</td>
</tr>
<tr>
<td>intercept[4][1]</td>
<td>21.782</td>
<td>10.251</td>
<td>0.754</td>
<td>22.287</td>
<td>38.032</td>
</tr>
<tr>
<td>intercept[5][1]</td>
<td>33.331</td>
<td>13.953</td>
<td>8.441</td>
<td>33.551</td>
<td>58.951</td>
</tr>
</tbody>
</table>

(table continues)
<table>
<thead>
<tr>
<th>Param</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>knot[1]</td>
<td>7.347</td>
<td>1.211</td>
<td>7.000</td>
<td>7.000</td>
<td>9.000</td>
</tr>
<tr>
<td>knot[2]</td>
<td>11.864</td>
<td>1.004</td>
<td>11.000</td>
<td>12.000</td>
<td>13.000</td>
</tr>
<tr>
<td>knot[3]</td>
<td>15.894</td>
<td>0.325</td>
<td>15.000</td>
<td>16.000</td>
<td>16.000</td>
</tr>
<tr>
<td>knot[4]</td>
<td>12.481</td>
<td>9.538</td>
<td>1.000</td>
<td>10.000</td>
<td>32.000</td>
</tr>
<tr>
<td>knot[5]</td>
<td>11.919</td>
<td>9.612</td>
<td>1.000</td>
<td>9.000</td>
<td>32.000</td>
</tr>
<tr>
<td>rho[1]</td>
<td>0.161</td>
<td>0.118</td>
<td>0.000</td>
<td>0.139</td>
<td>0.383</td>
</tr>
<tr>
<td>rho[2]</td>
<td>0.091</td>
<td>0.082</td>
<td>0.000</td>
<td>0.067</td>
<td>0.256</td>
</tr>
<tr>
<td>rho[3]</td>
<td>0.131</td>
<td>0.108</td>
<td>0.000</td>
<td>0.103</td>
<td>0.343</td>
</tr>
<tr>
<td>rho[4]</td>
<td>0.100</td>
<td>0.087</td>
<td>0.000</td>
<td>0.077</td>
<td>0.270</td>
</tr>
<tr>
<td>rho[5]</td>
<td>0.107</td>
<td>0.094</td>
<td>0.000</td>
<td>0.080</td>
<td>0.294</td>
</tr>
<tr>
<td>slope1[1]</td>
<td>-0.812</td>
<td>1.431</td>
<td>-4.148</td>
<td>-0.580</td>
<td>1.664</td>
</tr>
<tr>
<td>slope1[2]</td>
<td>0.970</td>
<td>0.714</td>
<td>-0.346</td>
<td>0.932</td>
<td>2.400</td>
</tr>
<tr>
<td>slope1[3]</td>
<td>0.631</td>
<td>0.573</td>
<td>-0.502</td>
<td>0.630</td>
<td>1.764</td>
</tr>
<tr>
<td>slope1[4]</td>
<td>0.074</td>
<td>1.750</td>
<td>-3.100</td>
<td>-0.124</td>
<td>3.148</td>
</tr>
<tr>
<td>slope1[5]</td>
<td>0.513</td>
<td>1.757</td>
<td>-2.131</td>
<td>0.175</td>
<td>4.233</td>
</tr>
<tr>
<td>slope2[1]</td>
<td>-0.190</td>
<td>0.237</td>
<td>-0.632</td>
<td>-0.185</td>
<td>0.284</td>
</tr>
<tr>
<td>slope2[2]</td>
<td>-0.418</td>
<td>0.268</td>
<td>-0.935</td>
<td>-0.419</td>
<td>0.130</td>
</tr>
<tr>
<td>slope2[3]</td>
<td>-0.361</td>
<td>0.443</td>
<td>-1.266</td>
<td>-0.356</td>
<td>0.478</td>
</tr>
<tr>
<td>slope2[4]</td>
<td>-0.740</td>
<td>0.865</td>
<td>-1.761</td>
<td>-0.870</td>
<td>0.830</td>
</tr>
<tr>
<td>slope2[5]</td>
<td>-0.479</td>
<td>0.956</td>
<td>-1.658</td>
<td>-0.562</td>
<td>1.215</td>
</tr>
</tbody>
</table>
Figure 4. Trace plots for the monitored parameters (four chains) for the multiple baseline design.
Figure 4 (cont.). Trace plots for the monitored parameters (four chains) for the multiple baseline design.
Figure 5. Density plots for the monitored parameters (four chains) for the multiple baseline design.
Figure 5 (cont.). Density plots for the monitored parameters (four chains) for the multiple baseline design.
The density and trace plots for four chains indicate convergence of parameter estimates. Based on the estimates of the posterior distributions of the parameters in Table 2 and Figure 5 we conclude as follows. The autocorrelation parameters \( \rho[i], \quad i=1, 2, 3, 4, 5 \) indicate that the data are serially correlated. Tier 1 has the highest autocorrelation \( \rho[1] \). Tier 2 has the highest stability with the lowest standard deviation \( \sigma[2] \). Tier 5 has the lowest stability.

The level changes between the baseline and intervention phases for the first three tiers are statistically significant \( \text{dint}[i], \quad i=1, 2, 3 \). For the 3rd and 4th tiers, the 95% HDIs of the posterior distributions for the level change parameters include 0, indicating that the level changes are not statistically significant. Tier 3 has the largest level change \( \text{dint}[3] \) with the highest baseline level \( \text{int1}[3] = 58.035 \). Slopes and slope changes, however, were not statistically significant according to the 95% HDI of \( \text{slope}[i] \) and \( \text{dslope}[i], \quad i=1, 2, 3, 4, 5 \). This indicates flat series for all the 5 tiers with no evidence of slope changes.

According to the original report and interpretation by Rodriguez and Anderson (2014), the first 4 tiers showed immediate level changes at the change-points. The reduction from baseline to intervention was 78%, 80%, and 82% for the first 3 tiers, respectively. The change-point and level change estimates by the proposed model in the present study confirmed the effect of the intervention on tiers 1 and 3. For tier 2, the level change was confirmed, but the change-point estimates showed one time point delay of the intervention effect. The 65% reduction for the 4th tier and the small reduction for the 5th tier, however, were not supported by the estimates of the proposed
Data Analysis for Alternating Treatments Designs

DeVeney et al. (2014) compared the amount of words that late-talking toddlers learned from dense and sparse neighborhoods conditions. Participants received 3 baseline sessions and 8 treatment sessions. In each session, both dense and sparse conditions were applied. We demonstrate data analysis at the last session of the intervention phase. The data are presented in Figure 6.

Figure 6. Plot for the alternating treatments design data from DeVeney, Cress, and Reid, 2014.

The total numbers of different words used were zero during the entire baseline phase. When the intervention phase started, data from both the dense and sparse words conditions showed immediate increase. The trends for both conditions during intervention phase were positive. The dense words condition showed higher level than the sparse words condition, indicating that the dense words condition has better effect on increasing the total number of different words used.
The estimation time for the statistical model was 1 minutes. The mode of the posterior distributions of the change-point parameter is 4. The model accurately detected the change-point of the alternating treatments design. Table 3 presents the output summary of the estimation for the parameters: int1 and slope1, baseline level and trend; int2[i] and slope2[i] (i = 1, 2), intervention phase intercepts and slopes for two conditions; knot, estimated change-point; dint[i] and dslope[i] (i = 1, 2), level and slope changes across phases for two conditions; dintp and dslopep, level and slope difference across treatment conditions (dintp = int2[1] - int2[2], dslopep = slope2[1] - slope2[2]); rho[i] and sigma[i] (i = 1, 2), autocorrelation and standard deviation for each condition.

The trace plots and posterior density plots are shown in Figures 7 and 8.

Table 3

<table>
<thead>
<tr>
<th>Param.</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>dintercept[1]</td>
<td>2.529</td>
<td>0.701</td>
<td>1.166</td>
<td>2.631</td>
<td>3.724</td>
</tr>
<tr>
<td>dintercept[2]</td>
<td>1.426</td>
<td>0.795</td>
<td>-0.258</td>
<td>1.513</td>
<td>2.854</td>
</tr>
<tr>
<td>Dinterceptp</td>
<td>0.995</td>
<td>0.662</td>
<td>-0.339</td>
<td>1.013</td>
<td>2.284</td>
</tr>
<tr>
<td>dslope[1]</td>
<td>-0.116</td>
<td>0.305</td>
<td>-0.731</td>
<td>-0.090</td>
<td>0.419</td>
</tr>
<tr>
<td>dslope[2]</td>
<td>-0.040</td>
<td>0.356</td>
<td>-0.762</td>
<td>-0.007</td>
<td>0.607</td>
</tr>
<tr>
<td>Dslopep</td>
<td>-0.116</td>
<td>0.160</td>
<td>-0.426</td>
<td>-0.119</td>
<td>0.200</td>
</tr>
<tr>
<td>intercept1</td>
<td>0.696</td>
<td>0.585</td>
<td>4.0e-6</td>
<td>0.560</td>
<td>1.774</td>
</tr>
<tr>
<td>intercept2[2]</td>
<td>2.229</td>
<td>0.540</td>
<td>1.138</td>
<td>2.223</td>
<td>3.341</td>
</tr>
<tr>
<td>Knot</td>
<td>4.014</td>
<td>0.268</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>rho[1]</td>
<td>0.219</td>
<td>0.164</td>
<td>1.2e-7</td>
<td>0.186</td>
<td>0.535</td>
</tr>
<tr>
<td>rho[2]</td>
<td>0.193</td>
<td>0.15</td>
<td>1.3e-6</td>
<td>0.160</td>
<td>0.488</td>
</tr>
<tr>
<td>sigma[1]</td>
<td>0.577</td>
<td>0.154</td>
<td>0.334</td>
<td>0.548</td>
<td>0.879</td>
</tr>
<tr>
<td>sigma[2]</td>
<td>0.699</td>
<td>0.182</td>
<td>0.414</td>
<td>0.666</td>
<td>1.068</td>
</tr>
<tr>
<td>slope1</td>
<td>0.260</td>
<td>0.289</td>
<td>-0.234</td>
<td>0.230</td>
<td>0.843</td>
</tr>
<tr>
<td>slope2[1]</td>
<td>0.144</td>
<td>0.101</td>
<td>-0.050</td>
<td>0.140</td>
<td>0.342</td>
</tr>
<tr>
<td>slope2[2]</td>
<td>0.260</td>
<td>0.126</td>
<td>0.004</td>
<td>0.261</td>
<td>0.508</td>
</tr>
</tbody>
</table>

Note. a Condition [1] is for the words from Dense neighborhoods. b Condition [2] is for the words from Sparse neighborhoods.
Figure 7. Trace plots for the monitored parameters (four chains) for the alternating treatments design.
Figure 8. Density plots for the monitored parameters (four chains) for the alternating treatments design.
The trace and density plots for four chains indicate convergence of parameter estimates. Based on the posterior distributions of the parameters estimates in Table 3 and Figure 8 we make the conclusions as follows. The autocorrelation parameters \( \rho[i], i=1, 2 \) indicate that the data are serially correlated. The data from the dense words condition has better stability \( \sigma[1] < \sigma[2] \), but higher autocorrelation \( \rho[1] > \rho[2] \). Level changes between the baseline and intervention phases for both conditions are statistically significant \( \text{dint}[i], i=1,2 \). Trends for both conditions in the intervention phase are positive \( \text{slope2}[i], i=1,2 \).

In the intervention phase, the level of the dense words condition \( \text{int2}[1] \) is higher than the level of the sparse words condition \( \text{int2}[2] \). As a result, level change between the two phases is larger under the dense words condition \( \text{dint}[1] \). However, the sparse words condition \( \text{slope2}[2] \) has larger slope than the dense words condition \( \text{slope2}[1] \).

When comparing the effects of the two treatments in the intervention phase, the estimate of the level difference of the two treatment conditions \( \text{dintp} \) is positive and statistically significantly different from 0. However, the estimates of the slope differences across the two treatment conditions \( \text{dslopep} \) are negative. The results indicate that both interventions are effective, but the dense words condition \( i=1 \) has larger effect on the level, yet the sparse words condition \( i=2 \) has larger effect on the trend.

Data Analysis for Changing Criterion Designs

Cameron et al. (2005) used a changing criterion design to examine the effect of positive behavioral interventions. The data plot is as shown in Figure 9. The plot indicates that the data follows a linear pattern. Therefore it is proper to use a linear
model for the data analysis. The models developed in this study are confirmatory models. They are not used for prediction purposes.

Figure 9. Plot for the changing criterion design data from Cameron et al. (2005).

The estimation time for the statistical model was 35.3 seconds. The output summary in Table 4 represents the estimation for the parameters: int (intercept), rho (autocorrelation), sigma (standard deviation), and slope. The trace plots and posterior density plots are shown in Figures 10 and 11.

Table 4

Bayesian Estimation Output Summary for the Changing Criterion Design Data Analysis

<table>
<thead>
<tr>
<th>Param</th>
<th>Mean</th>
<th>SD</th>
<th>Lower95</th>
<th>Median</th>
<th>Upper95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.906</td>
<td>3.162</td>
<td>0.002</td>
<td>4.484</td>
<td>10.760</td>
</tr>
<tr>
<td>rho</td>
<td>0.747</td>
<td>0.062</td>
<td>0.628</td>
<td>0.749</td>
<td>0.871</td>
</tr>
<tr>
<td>sigma</td>
<td>4.528</td>
<td>0.373</td>
<td>3.814</td>
<td>4.500</td>
<td>5.270</td>
</tr>
<tr>
<td>slope</td>
<td>1.060</td>
<td>0.073</td>
<td>0.915</td>
<td>1.064</td>
<td>1.200</td>
</tr>
</tbody>
</table>
Figure 10. Trace plots for the monitored parameters (four chains) for the changing criterion design.

Figure 11. Density plots for the monitored parameters (four chains) for the changing criterion design.
The trace and density plots for four chains in Figures 10 and 11 indicate convergence of parameter estimates. According to the output summary in Table 4, the linear procedure starts from a lower valued intercept (int) and increased with a positive trend (slope). The estimates of the slope parameter have narrow HDI that do not include zero. This confirms the intervention effect on enhancing the target behavior. The estimation for the serial dependency (rho, autocorrelation) and stability of the data (sigma, standard deviation) are also reported in Table 4. Based on the model equation in the JAGS code, the estimated values are calculated and plotted on the original graph, as shown in Figure 12.

![Figure 12](image)

**Figure 12.** Plot for the changing criterion design data with the estimated regression line (empty circle).

The estimated values (empty circles in Figure 11) do not fall on one straight line, but follow the similar pattern of the observed data with a slope. This is because the
autocorrelation is quite high (mean of \( \rho = 0.747 \)). The estimated values follow the average level and trend of the actual observations. This indicates high estimation quality of the proposed statistical model.

Discussion

The present study is an extension of the first manuscript of this dissertation. We have applied the change-point estimation approach to more designs in SSR. Three linear modeling approaches have been described to investigate Bayesian data analysis approaches for multiple baseline design, alternating treatments design (ATD), and changing criterion design. We address the problem of serial dependency; estimate the change-points as unknown parameters to discover the effect of the intervention; and present a Bayesian posterior distribution analysis procedure for parameter estimation. Current statistical methods for analysis of SSDs have applied Bayesian methods in the A-B and A-B-A-B designs. However, this is the first time that Bayesian analysis is applied to the other designs. Multiple baseline design is the most popular design in SSR. ATD and changing criterion design serve different purposes in the application of SSDs. The present study presents complete Bayesian data analysis procedures for these commonly used SSDs.

In social sciences, establishing causality is very important but difficult to achieve. In SSR, multiple baseline design provides better control for the internal validity threat of history. Problems of maturation, statistical regression, initial measurement effects (a.k.a. testing effects), and instrumentation may be controlled by achieving a stable baseline with flat trend. When the baseline phase shows a trend in the desired direction, it is more difficult to demonstrate the effectiveness of the intervention. In the multiple
baseline design and ATD, estimating change-points as unknown parameters provides additional evidence for causal effect of the intervention on target behavior. The proposed approaches detect (1) the existence of changes between phases, (2) the time points when the changes take place and, (3) the type of changing patterns on level and trend. Accurately estimating the value of the change-point confirms the changes between baseline and intervention phases. Therefore the proposed approach adds to evidence of causality. Change-point parameter with wide HDI, however, indicates little or no evidence of the intervention effect. The pattern of the change is described by the measures of level and trend changes. The proposed approaches provide evidence of the intervention effect on level and trend in detail.

In describing the change-point evaluation approaches, we have provided analysis methods for the multiple baseline design and ATD. We have defined approaches that can estimate change-points for all the tiers in the multiple baseline design and the one or two phase changes in the ATD. In this study we recommend that inference be based on a comprehensive analysis of the estimates of all the change-points.

The procedures that we have developed in this article are based on the linear regression model. We have addressed the problem of serial dependency by taking autocorrelation into account. The approach we suggest for the changing criterion design is a direct application of the linear regression model. This meets the demand of evaluating the effectiveness of the intervention. For the data that show a nonlinear pattern of the time-series, the linear equation should be replaced by appropriate polynomial function. This is an avenue for future research.
Another possible direction for future exploration is to quantify the sensitivity of the change-point detector. In the current model, the change-point parameter could only be detected accurately when the data pattern change is significant. When there is not much change at the point of phase change, for example in the third and fourth tiers of the multiple baseline design example, the proposed model failed to estimate the change-point parameters with enough accuracy. A simulation study could be conducted to examine the sensitivity of the model on the change-point parameter.

In the first manuscript of this dissertation, we have explored data analysis approaches for the baseline phase, A-B design and A-B-A-B design. Together with the three models in the second study, the dissertation has presented a set of six statistical models for the data analysis of SSR. The detailed information provided by the model estimation has offered single subject researchers a series of statistical models to analyze the results and draw inferences about the intervention effect.
References


Appendix D

R and JAGS Codes for the Multiple Baseline Design
JAGS Code (Multiple Baseline.txt):

model{
  for (i in 1:I){
    Yhat[i,1] <- slope1[i] + (int1[i] - slope1[i] * knot[i])
    Y[i,1] ~ dnorm(Yhat[i,1], tau.e[i])
    D[i,1] <- 0
    for (t in 2:T){
      D[i,t] <- step(t - knot[i])
      Yhat[i,t] <- ((1 - D[i,t]) - rho[i] * (1 - D[i,(t-1)])) * (int1[i] - slope1[i] * knot[i]) +
                    ((1 - D[i,t]) * t - rho[i] * (1 - D[i,(t-1)]) * (t-1)) * slope1[i] +
                    (D[i,t] - rho[i] * D[i,t-1]) * (int2[i] - slope2[i] * knot[i]) +
                    (D[i,t] * t - rho[i] * D[i,(t-1)]) * (t-1)) * slope2[i] + rho[i] * Y[i,(t-1)]
      Y[i,t] ~ dnorm(Yhat[i,t], tau.delta[i])
    }
    tau.e[i] <- 1/pow(sigma.e[i],2)
    sigma[i] <- 1/sqrt(tau.delta[i])
    sigma.e[i] <- sigma[i]/sqrt(1 - pow(rho[i],2))
    dint[i] <- int2[i] - int1[i]
    dslope[i] <- slope2[i] - slope1[i]
    int1[i] ~ dunif(0,ymax)
    int2[i] ~ dunif(0,ymax)
    slope1[i] ~ dnorm(0,precs1[i])
    slope2[i] ~ dnorm(0,precs2[i])
    precs1[i] ~ dgamma(1,1)
    precs2[i] ~ dgamma(1,1)
    for (t in 1:T){
      p[i,t] <- 1/T
    }
    knot[i] ~ dcat(p[i,])
    tau.delta[i] ~ dgamma(1,1)
    rho[i] ~ dbeta(1,5)
  }
  #data# I, T, Y, ymax
}
R Code (Multiple Baseline.R):

install.packages('rjags')                           #Install and load R packages
install.packages('runjags')
install.packages('prettyR')
require(rjags)
require(runjags)
require(prettyR)
setwd('/Volumes/LEXAR/Dissertation/Data Analysis/Multiple Baseline')          #Set working directory
Y <- as.matrix(read.csv('Rodriguez 2014.csv', header=F))   #Load data file
I <- nrow(Y)                                       #Number of tiers
T <- ncol(Y)
ymax <- max(Y, na.rm = T) #maxima of the input values, missing values removed
#plot the data
y1 <- as.numeric(Y[1,])
y2 <- as.numeric(Y[2,])
y3 <- as.numeric(Y[3,])
y4 <- as.numeric(Y[4,])
y5 <- as.numeric(Y[5,])
par(mfrow=c(5,1)) #Place graphs on the page in 3 rows and 1 column
par(mar=c(2.5, 4.5, 2, 1)) #Change the margins of the actual graphs
#plot the 1st tier
plot(y1, type="o", bty="l", pch=16, ylim=c(0,ymax), ylab="", cex.lab=1.5)
abline(v=6.5, lty=3)
mtext("Baseline", side=3, line=0, outer=F, adj=0, cex=0.9)
mtext("TGC Intervention", side=3, line=0, outer=F, adj=0.33, cex=0.9)
text(27, 50, "Deborah's Group")
#Add the 2nd tier
plot(y2, type="o", bty="l", pch=16, ylim=c(0,ymax), ylab="with Problem Behavior", cex.lab=1.2)
abline(v=10.5, lty=3)
text(27, 50, "Amy's Group")
#Add the 3rd tier
plot(y3, type="o", bty="l", pch=16, ylim=c(0,ymax), ylab="Intervals Scored", cex.lab=1.2)
abline(v=15.5, lty=3)
text(27, 50, "Barbara's Group")
#Add the 4th tier
plot(y4, type="o", bty="l", pch=16, ylim=c(0,ymax), ylab="Percent of 10-Second", cex.lab=1.2)
abline(v=19.5, lty=3)
text(27, 50, "Natasha's Group")
#Add the 5th tier
plot(y5, type="o", bty="l", pch=16, ylim=c(0,ymax), ylab="", cex.lab=1.5)
abline(v=19.5, lty=3)
text(27, 50, "Candice's Group")
#run the analysis
runjags.object <- autorun.jags("Multiple Baseline.txt",
    monitor = c("slope1", "slope2", "int1", "int2",
              "dslope", "dint",
              "knot", "rho", "sigma"),
    n.chains=4, startburnin = 25000, startsample = 4000,
thin.sample = FALSE, thin = 1)

#mode for the change-point
runjags.object                                      #Output Summary
knot1.est <- as.mcmc.list(runjags.object,vars="knot[1]"
runjags.object
knot2.est <- as.mcmc.list(runjags.object,vars="knot[2]"
runjags.object
knot3.est <- as.mcmc.list(runjags.object,vars="knot[3]"
runjags.object
knot4.est <- as.mcmc.list(runjags.object,vars="knot[4]"
runjags.object
knot5.est <- as.mcmc.list(runjags.object,vars="knot[5]"
runjags.object
knot1.est <- as.matrix(knot1.est)
knot2.est <- as.matrix(knot2.est)
knot3.est <- as.matrix(knot3.est)
knot4.est <- as.matrix(knot4.est)
knot5.est <- as.matrix(knot5.est)
knot1.mode
knot2.mode
knot3.mode
knot4.mode
knot5.mode
plot(runjags.object, type="trace", layout=c(8,3))
plot(runjags.object, type="density", layout=c(8,3))

Data file (Rodriguez 2014.csv):

<p>| | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

85
Appendix E

R and JAGS Codes for the Alternating Treatments Design
JAGS Code for two phases design (ATD2phases.txt):

```jag`
model{
  D[1] <- 0

  for (t in 2:T){
    D[t] <- step(t - knot)
  }

  for (t in 1:T){
    p[t] <- 1/T
  }

  knot ~ dcat(p[])

  for (i in 1:I){
    Yhat[i,1] <- slope1[i] + (int1[i] - slope1[i] * knot)
    Y[i,1] ~ dnorm(Yhat[i,1], tau.e[i])

    for (t in 2:T){
      Yhat[i,t] <- ((1 - D[t]) - rho[i] * (1 - D[t-1])) * (int1[i] - slope1[i] * knot) +
      ((1 - D[t]) * t - rho[i] * (1 - D[t-1]) * (t-1)) * slope1[i] +
      (D[t] - rho[i] * D[t-1]) * (int2[i] - slope2[i] * knot) +
      (D[t] * t - rho[i] * D[t-1] * (t-1)) * slope2[i] + rho[i] * Y[i,(t-1)]
      Y[i,t] ~ dnorm(Yhat[i,t], tau.delta[i])
    }

    tau.e[i] <- 1/pow(sigma.e[i],2)
    sigma[i] <- 1/sqrt(tau.delta[i])
    sigma.e[i] <- sigma[i]/sqrt(1 - pow(rho[i],2))
    dint[i] <- int2[i] - int1[i]
    dslope[i] <- slope2[i] - slope1[i]

    int1[i] ~ dunif(0,ymax)
    int2[i] ~ dunif(0,ymax)
    slope1[i] ~ dnorm(0,precs1[i])
    slope2[i] ~ dnorm(0,precs2[i])
    precs1[i] ~ dgamma(1,1)
    precs2[i] ~ dgamma(1,1)
    tau.delta[i] ~ dgamma(1,1)
    rho[i] ~ dbeta(1,5)
  }

  #data# I, T, Y, ymax
}
```
R Code for two phases design (ATD2phases.R):

install.packages('rjags')  #Install and load R packages
install.packages('runjags')
install.packages('prettyR')
require(rjags)
require(runjags)
require(prettyR)
setwd('E:/Dissertation/Data Analysis/ATD')  #Set working directory
Y <- as.matrix(read.csv('DeVeney et al 2014.csv', header=F))  #Load data file
I <- nrow(Y)  #Number of treatments
T <- ncol(Y)
ymax <- max(Y, na.rm = T)  #maxima of the input values, missing values removed
#plot the data
y1 <- as.numeric(Y[1,])
y2 <- as.numeric(Y[2,])
plot(y1, type="o", bty="l", pch=16, ylim=c(0,ymax), cex.lab=0.8,
   xlab="Experimental Sessions", ylab="Total# of Different Words Used")
lines(y2, type="o", pch=0)
abline(v = 3.5, lty = 1)

#run the analysis
runjags.object <- autorun.jags("ATD2phases.txt",
   monitor = c("slope1[1]", "slope2", "dslope", "dslopep",
   "int1[1]", "int2", "dint","dintp",
   "knot", "rho", "sigma"),
   n.chains=4, startburnin = 25000, startsample = 4000,
   thin.sample = FALSE, thin = 1)
runjags.object  #Output Summary
#mode for the change-point
knot.est <- as.mcmc.list(runjags.object,vars="knot")
knot.est <- as.matrix(knot.est)
knot.mode <- Mode(knot.est)
plot(runjags.object, type="trace", layout=c(3,3))
plot(runjags.object, type="density", layout=c(3,3))
JAGS Code for three phases design (ATD3phases.txt):

model{
    D1[1] <- 0
    D2[1] <- 0
    for (t in 2:T){
        D1[t] <- step(t - knot1)
        D2[t] <- step(t - knot2)
    }
    knot1 ~ dunif(1, T)
    knot2 ~ dunif(1, T)

    for (i in 1:I){
        Yhat[i,1] <- slope1[i] + (int1[i] - slope1[i] * knot1)
        Y[i,1] ~ dnorm(Yhat[i,1], tau.e[i])

        for (t in 2:T){
            Yhat[i,t] <- ((1 - D1[t]) - rho[i] * (1 - D1[t-1])) * (int1[i] - slope1[i] * knot1)
                + ((1 - D1[t]) * t - rho[i] * (1 - D1[t-1]) * (t - 1)) * slope1[i]
                + (D1[t] * (1 - D2[t]) - rho[i] * D1[t-1] * (1 - D2[t-1])) * (int2[i] - slope2[i] * knot1)
                + (D1[t] * (1 - D2[t]) * t - rho[i] * D1[t-1] * (1 - D2[t-1]) * (t - 1)) * slope2[i]
                + (D1[t] * D2[t] - rho[i] * D1[t-1] * D2[t-1]) * (int3[i] - slope3[i] * knot2)
                + (D1[t] * D2[t] * t - rho[i] * D1[t-1] * D2[t-1] * (t - 1)) * slope3[i]
                + rho[i] * Y[i,t-1]
            Y[i,t] ~ dnorm(Yhat[i,t], tau.delta[i])
        }
    }

    tau.e[i] <- 1/pow(sigma.e[i],2)
    tau.delta[i] <- 1/pow(sigma[i],2)
    sigma.e[i] <- sigma[i]/sqrt(1 - pow(rho[i],2))

    int1[i] ~ dunif(0,ymax)
    int2[i] ~ dunif(0,ymax)
    int3[i] ~ dunif(0,ymax)
    slope1[i] ~ dnorm(0,precs1[i])
    slope2[i] ~ dnorm(0,precs2[i])
    slope3[i] ~ dnorm(0,precs3[i])
    precs1[i] ~ dgamma(1,1)
    precs2[i] ~ dgamma(1,1)
    precs3[i] ~ dgamma(1,1)
    sigma[i] ~ dunif(0,10)
    rho[i] ~ dbeta(1,5)
}
#data# I, T, Y, ymax
R Code for three phases design (ATD3phases.R):

```r
# ATD with Best alone phase (phase III)
install.packages('rjags')  # Install and load R packages
install.packages('runjags')
require(rjags)
require(runjags)
setwd('E:/Dissertation/Codes/Appendix')  # Set working directory
Y <- as.matrix(read.csv('ATDfake2.csv', header=F))  # Load and specify data file
I <- nrow(Y)  # Number of treatments
T <- ncol(Y)
ymax <- max(Y, na.rm = T)  # maxima of the input values, missing values removed
runjags.object <- autorun.jags("ATD3phases.txt", monitor = c("slope1", "slope2",
                  "slope3", "int1", "int2", "int3", "knot1", "knot2", "rho", "sigma"), n.chains=2,
                  startburnin = 25000, startsample = 4000, thin.sample = FALSE, thin = 1)
runjags.object  # Output Summary
plot(runjags.object, type="trace", layout=c(2,2))
plot(runjags.object, type="density", layout=c(2,2))
```

Data file (DeVeney et al 2014.csv):

```
0  0  0  3  3  4  4  4  4  4  4  4
0  0  0  2  2  3  3  4  4  3  4
```

Data file (ATDfake2.csv):

```
2  8  2  2  2  8  2  9  2  8  3  0  2  9  2  8  9  2  0  0  1  1  1  0  1  1  2  1  1  2  1  1
2  1  2  1  2  2  2  0  2  2
```
Appendix F

R and JAGS Codes for the Changing Criterion Design
JAGS Code (Changing Criterion.txt):

model
{
    Yhat[1] <- slope + int
    Y[1] ~ dnorm(Yhat[1], tau.e)
    tau.e <- 1/pow(sigma.e,2)

    for (t in 2:T){
        Yhat[t] <- (1-rho)*int + (t-rho*(t-1))*slope + rho*Y[t-1]
        Y[t] ~ dnorm(Yhat[t], tau.delta)
    }
    sigma <- 1/sqrt(tau.delta)

    sigma.e <- sigma/sqrt(1-pow(rho,2))
    int ~ dunif(0,ymax)
    slope ~ dnorm(0,precs[1])
    precs[1] ~ dgamma(1,1)
    rho ~ dbeta(1,5)
    tau.delta ~ dgamma(1,1)
    #data# T, Y, ymax
}

R Code (Changing Criterion.R):

install.packages('rjags')                           #Install and load R packages
install.packages('runjags')
require(rjags)
require(runjags)
setwd('D:/Dissertation/Data Analysis/Changing Criterion')    #Set working directory
data <- as.matrix(read.csv('Cameron et al 2005.csv', header=F))   #Load data file
Y <- as.numeric(data[,1])
T <- ncol(data)
ymax <- max(Y, na.rm = T)     #maxima of the input values, missing values removed
#plot the data
plot(Y, type="o", bty="l", pch=20, ylim=c(0,ymax), cex.lab=0.8, xlab="Sessions",
     ylab="% of Steps Completed")
abline(v = 1.5, lty = 1)
abline(v = 12.5, lty = 3)
abline(v = 24.5, lty = 3)
abline(v = 33.5, lty = 3)
abline(v = 40.5, lty = 3)
abline(v = 70.5, lty = 3)
abline(v = 73.5, lty = 3)
abline(v = 76.5, lty = 3)

#Run the analysis, monitoring slope/intercept/standard deviation/autocorrelation
runjags.object <- autorun.jags("Changing Criterion.txt", monitor = c("int", "slope", "sigma", "rho"),
  n.chains=4, startburnin = 25000, startsample = 4000,
  thin.sample = FALSE, thin = 1)
runjags.object                                      #Output Summary
plot(runjags.object, type="trace", layout=c(2,2))
plot(runjags.object, type="density", layout=c(2,2))

#add regression line
int = 4.9058
rho = 0.74696
slope = 1.0604
Yhat = numeric(T)
Yhat[1] = slope + int
for (t in 2:(T)){
  Yhat[t]= int * (1 - rho) + slope * (t - rho * (t - 1)) + rho * Y[t-1]
}
lines(Yhat, type="o", pch=1)

Data file (Cameron et al 2005.csv):

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>12</th>
<th>0</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>75</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>88</td>
<td>88</td>
<td>88</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
EXTENDED LITERATURE REVIEW

The two commonly used research designs to evaluate intervention effects are group-comparison experimental designs and single-case experimental designs (SCED, a.k.a. single subject research designs; Barlow & Hersen, 1984; Franklin, Allison, & Gorman, 1997). In certain research domains and occasions when the population is small or highly fragmented, or where random assignment is practically or ethically impossible, SCED has become very popular (Van den Noortgate & Onghena, 2003). Since the 1990s, single subject research (SSR) has been widely applied in psychology, education (Franklin et al., 1997), medicine (Zucker, Schimid, McIntosh, Agostino, Selker, & Lau, 1997), and school psychology (Kratochwill, 1985; Wacker, Steege, & Berg, 1988).

Compared to group-comparison designs, SCED demand fewer participants and yield relatively low costs and immediate valuable findings (Barlow, Nock, & Hersen, 2009). Single subject researchers intensively study individual participant’s behavior. They explore the effect of the intervention by comparing participant’s behavior in different phases (baseline phase without intervention, intervention phase, etc.). SCED emphasize individualized scores and verify the intervention effects through replication. Single subject designs (SSD) have many names: intra-subject replication, intra-subject design, reversal design, individual organism research, n-of-1 design, n of 1 data, n=1 study, and one-subject experiment (Shadish & Sullivan, 2011). Different types of designs in SSR vary from withdrawal and reversal designs (e.g. A-B design, A-B-A design, A-B-A-B design and its variations), multiple baseline and multiple probe designs, comparative intervention designs (e.g. alternating treatments design, multi-
element design, and parallel-treatments design), to combination designs (Gast, 2010).
In the data analysis procedure, SSR uses both visual and statistical analysis to
determine the functional relationship between the independent and dependent variables
(Gast, 2010).

Research Designs in SSR
A-B Design

A-B design is the most basic single subject design (Riley-Tillman & Burns, 2009).
It features one baseline phase (phase A) and one intervention phase (phase B). The
baseline phase contains raw data without any intervention or treatment. Once the
baseline data shows stability, intervention will be introduced and the intervention phase
starts. Figure 1 demonstrates the basic A-B design.

In Figure 1, the target behavior is a certain type of problem behavior. The
purpose was to reduce the problem behavior. Baseline data collection started from day
one and lasted till day fifteen. During the baseline phase, the problem behavior showed
stability. Therefore it was proper to start to implement the intervention, assuming the
intervention will reduce the problem behavior. On day sixteen, researcher introduced
the intervention into the experiment. Compared to the problem behavior in day fifteen,
day sixteen showed a big jump downward. Through the intervention phase, the problem
behavior stayed at a very low level. The figure shows the intervention effect directly and
intuitively.

In this research design, the independent variable was the intervention and the
dependent variable was the problem behavior. A-B design is valid when the research
environment is well controlled and potential spurious variables do not influence the change of the dependent variable. However, A-B design does not test the replication effect.

A-B-A-B Design

A-B-A-B design is one type of withdrawal and reversal design. It had become one of the most commonly used single subject designs in behavioral science (Gast & Hammond, 2010). Following the intervention phase in the A-B design, A-B-A-B design adds another baseline conditioned phase and another intervention phase. As the baseline and intervention phase named phase A and phase B in the A-B design, the four phases in A-B-A-B design are called phase A₁, B₁, A₂, and B₂. Phase A₁ and A₂ are baseline conditioned phases. Phase B₁ and B₂ are intervention phases. From phase B₁ to A₂, the intervention is withdrawn. This requires the target behavior to be reversible. In the last phase (B₂), the intervention is applied again. Figure 2 demonstrates the basic A-B-A-B design.

INSERT FIGURE 2 ABOUT HERE

The most important feature of this design is that it provides the evaluation on the replication effect (Gast & Hammond, 2010). Swaminathan, Roger, and Horner (2014) evaluated a single subject research that used A-B-A-B design with multiple participants. The formulas for the linear model with serial dependency can be found in the later section (equations 4, 5, 6, and 7).

Other Commonly Used SSR Designs

Multiple baseline design is very popular in SSR (Shadish & Sullivan, 2011). It evaluates replication effect by varies subjects, conditions, or behaviors. Multiple
baseline design collects data across three or more tiers, which can be applied to different behaviors, conditions, or participants (Gast & Ledford, 2010a). In each tier, the model is similar to the A-B design. Different tiers start the baseline at the same time, but introduce the intervention at different time points. Figure 3 demonstrates the basic multiple baseline design.

INSERT FIGURE 3 ABOUT HERE

Alternating treatments design (ATD) can be used to compare multiple interventions or conditions. Similar to the A-B-A-B design, ATD also requires the behavior to be reversible. In ATD, the baseline phase is recommended but not mandatory, so does the best alone phase (phase III), where only the superior intervention method is remained (Wolery, Gast, & Hammond, 2010). Figure 4 demonstrates the basic ATD design.

INSERT FIGURE 4 ABOUT HERE

Changing criterion design is used for instructional behaviors. It is treated as a variation of the multiple baseline design (Gast & Ledford, 2010b). Following a baseline which had shown stability, a series of criterion changes were applied onto the target behavior. Figure 5 used the teaching example from Gast and Ledford (2010b) to demonstrate the changing criterion design.

INSERT FIGURE 5 ABOUT HERE

Data Analysis in SSR

Visual analysis and statistical analysis are the two categories in the academic practice of single subject research (Gast, 2010; Houle, 2009). Each of the two analysis methods contributes to the data interpretation from different aspects. Visual analysis is
simple, graphical, and direct. Statistical analysis is objective and provides quantified evaluation of the intervention effect. Since 1990th, the controversy between these two methods has never stopped (Houle, 2009).

In single subject designs, the techniques of visual analysis have been very well developed and widely applied (Gast, 2010). Most of the SSR data have been analyzed through visual inspection (Barlow & Hersen, 1984; Campbell & Herzinger, 2010). Visual analysis is the most direct way to demonstrate the results of SSR and to communicate between scholars (Houle, 2009). In SSR, data graphing and visual analysis started since the beginning of baseline data collection. Baseline data are recorded and plotted repeatedly every day. Raw data are analyzed frequently (Gast & Spriggs, 2010). However, using only the visual analysis is not sufficient. Visual analysis is less reliable and more likely to yield to Type I errors (Kazdin, 1984). Visual analysis is lack of universal decision rules for the interpretation of the intervention effect (Campbell & Herzinger, 2010). Different researchers may explain the output differently for the same graphic result, especially when the data are relatively unstable (DeProspero & Cohen, 1979; Campbell & Herzinger, 2010).

Statistical analysis can be used to identify smaller effects and help increase confidence in visual analysis (Gorman & Allison, 1997). Compared to visual analysis, statistical methods are more objective, because they provide quantified output to reflect the effect of the intervention (Campbell & Herzinger, 2010). Recent advances of statistical analysis used various research methods to investigate and develop the data analysis methods in SSR. Those methods included effect size calculation (Parker, Hagan-Burke, & Vannest, 2007; Campbell, 2004; Busk & Serlin, 1992; de Vries &

More specifically, Swaminathan et al. (2014) proposed an alternative effect size measure with slope and intercept changes in the model. Parker et al. (2007) proposed a new effect size index, the percentage of all non-overlapping data (PAND). de Vries and Morey (2013) added Bayes factor analysis to the A-B design effect size evaluation. Maggin, Swaminathan, Rogers, O'Keeffe, Sugai, and Horner (2011) proposed a new method for effect sizes in SSD. Shadish and Sullivan (2011) conducted a meta-analysis of 809 single-case designs, coded by type of design, number of cases and outcomes, data points, phases per case, and autocorrelations.

Van den Noortgate and Onghena (2003) introduced a method to combine SSR using hierarchical linear models. They pointed out that the traditional meta-analysis methods yield relatively sparse results on SSR. Researchers later developed multilevel linear modeling methods to combine single-case experimental data from different studies (Moeyaert et al., 2013a; Owens & Ferron, 2012; Moeyaert et al., 2013b). Owens and Ferron (2012) examined the fixed effects (overall average baseline level and treatment effect) and variance components of a three-level model using Monte Carlo methods. Moeyaert et al. (2013a) conducted a simulation study to model external
events in a multilevel SSD, followed by an empirical experiment as the validation (Moeyaert et al., 2013b). Moeyaert et al. (2013b) found that it was better to use a homogeneous set of studies and at least 30 studies to do the meta-analysis, however, the number of measurements and cases is not as important.

The methodology studies of statistical analysis indicated two aspects to be taken into account when applying statistical methods onto SSR data. Firstly, SSR data contains successive observations and consequently, tends to have the issue of serial dependency (Kazdin, 1984). Ignoring the existence of serial dependency may lead to misinterpretation of the intervention effect. To address this issue, the SSR data analyses need to include autocorrelation in the statistical models (Kazdin, 1984; Manolov, Solanas, Sierra, & Evans, 2011; Swaminathan et al., 2014). The technical details of serial dependency and autocorrelation will be discussed in the following section. Secondly, according to Barlow, Nock and Hersen (2009), time series data analysis requires at least 50 observations, or more adequately 50 observations per phase. In SSR practice, following a single participant for 50 or more time points is not very common (Shadish & Sullivan, 2011). How can the issue of serial dependency be addressed with limited observations (less than 50)? Discussion on Bayesian methods in a later section will help us answer this question.

Technical Details

Effect Size

Effect size analyses in SSD have long been applied in the single subject research practice. Busk and Serlin (1992) presented three effect size measurement approaches for A-B design as follows.
(a) The first approach used an analog of Glass’ delta, treating baseline as the control group. It divided the difference of the phase averages by the baseline standard deviation, as shown in equation 1.

(b) The second approach was developed from the first method. It adjusted the denominator into the pooled within-phase variances (equation 2).

(c) Based on the method used in the second approach, the third method calculated the overall effect size for multiple cases and multiple studies. Interested readers may refer to the original work for details.

\[
Effect \ size (\Delta) = \frac{M_B - M_I}{SD_B} \tag{1}
\]

\[
Effect \ size = \frac{M_B - M_I}{SD_{pooled}} \tag{2}
\]

\(M_B\) is the mean of baseline observations, \(M_I\) is the mean of intervention phase data, \(SD_B\) is the baseline standard deviation, and \(SD_{pooled}\) is the pooled within-phase standard deviation.

There are also other types of effect size measures in SSR. Campbell (2004) demonstrated and compared the performance of four types of effect sizes in SSR: mean baseline reduction (MBLR), percentage of nonoverlapping data (PND), percentage of zero data points (PZD), and a regression-based \(d\) statistic. Based on the PND effect size calculation, Parker et al. (2007) introduced a new effect size index, the percentage of all non-overlapping data (PAND), which is related to Pearson’s \(Phi\). As another development, Manolov and Solanas (2009) presented the percentage of nonoverlapping corrected data (PNCD). In 2011, based on a thorough literature review on SSR effect
sizes, Maggin et al. introduced a new method for calculating effect size using general least squares to model autocorrelation of the time series data.

Recent studies also developed the effect size calculation for multiple subjects, either for one study (Swaminathan et al., 2014), or for multiple studies in the meta-analysis (van den Noortgate & Onghena, 2003). However, not all of these studies considered serial dependency in the effect size calculation. In 2014, Swaminathan et al. presented the equation of effect size measure that modeled serial dependency. The magnitude of the effect size was the average of the differences of the predicted and observed values through the intervention phase (phase B), as shown in equation 3.

\[
\delta_{AB} = \beta_2 + \beta_3 \frac{n_B - 1}{2}
\]

(3)

\(\delta_{AB}\) is the effect size reflecting the behavior change between phase A and phase B, \(\beta_2\) is the intercept change between the two phases, \(\beta_3\) is the slope change between the two phases, and \(n_B\) is the number of observations in phase B. This equation can be used for the A-B design data with one subject.

In 2013, de Vries and Morey introduced a set of Bayes factor tests for single subject data with two phases (A-B design). In Bayesian methods, Bayes factor estimates the ratio of probability of the null hypothesis and the alternative hypothesis (Kruschke, 2011). When the Bayes factor is greater than 1, the null hypothesis has a higher probability to be true than the alternative hypothesis, indicating there is no intervention effect. When the Bayes factor is smaller than 1, the alternative hypothesis is favored and the intervention effect proved. Closer the Bayes factor to 0, larger the intervention effect. Bayes factor will be discussed in more detail in a later section.
Stability, Serial Dependency, and Autocorrelation

In single subject research data analysis, the stability of time series data is the basic of the data analysis procedure (Gast & Spriggs, 2010). One criterion of stability in visual analysis is to obtain three data points with the same value (Kennedy, 2005). Another commonly used criterion for level stability is the 80% of the data fall into 20% range of the median method (Gast & Spriggs, 2010). The stability of the data, however, has not been the focus of any statistical analysis method.

As discussed in an earlier section, the successive observations in SSR data tend to have serial dependency. In statistical significance testing (SST), one critical assumption is the independence of error components. The expected value of the error correlation is assumed to be zero ($r_{e_1e_2} = 0$). However, in continuous or repeated measures over time, the assumption of independence-of-observations is usually not met (Kazdin, 1984). When the autocorrelation of error terms does not equal to zero, this type of data with successive observations have serial dependency. In interrupted time series analysis (ITSA), as will be discussed in the following section, the problem of serial dependency is solved by adding autocorrelation in the model.

In a time series, autocorrelation refers to the correlation between the current data points and the previous data points one or more time intervals (lags) ago. In social science, the autocorrelation parameter (number of lags) rarely exceed to one (McDowall, McCleary, Meidinger, & Hay, 1980). To discover serial dependency in SSR data, autocorrelation of lag 1 is sufficient (Kazdin, 1984). Swaminathan et al. (2014) presented the equations of the regression model and the serial dependency of the residual ($e_t$).
\[ y_t = \beta_0 + \beta_1 t + e_t \quad (4) \]
\[ e_t = \rho e_{t-1} + \epsilon_t \quad (5) \]

\( y_t \) is the predicted value of the target behavior at time \( t \); \( \beta_0 \) and \( \beta_1 \) are the intercept and slope coefficients of the linear regression model; \( e_t \) is the error of the regression model at time \( t \); \( \rho \) is the autocorrelation coefficient; and \( \epsilon_t \) is the independently distributed error. In time series analysis, this error is called the white noise (McDowall et al., 1980).

By putting these two equations together and doing some transformation\(^3\), the regression model with serial dependency becomes:

\[ y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1[t - \rho(t - 1)] + \epsilon_t \quad (6) \]

or

\[ y_t^* = \beta_0^* + \beta_1^* t^* + \epsilon_t, t = 2, 3, ..., n_t \quad (7) \]

Where \( y_t^* = y_t - \rho y_{t-1}, \beta_0^* = \beta_0(1 - \rho), \) and \( t^* = t - \rho(t - 1) \).

Interrupted Time Series Analysis (ITSA)

Time series quasi-experiments assess the impact of a discrete intervention on a social process (McDowall et al., 1980). This type of experiments has been widely applied in the area of legal impact assessment. Experimental psychologists have also been using them to test and measure the impacts of treatments (Gottman & McFall, 1972; Hall, Fox, Willard, Goldsmith, Emerson, Owen, Davis, & Porcia, 1971; Tyler & Brown, 1968). Examples of time series quasi-experiments have two things in

\(^3\) For those interested in greater detail:

\[ y_t = \beta_0 + \beta_1 t + e_t, \quad e_t = \rho e_{t-1} + \epsilon_t \Rightarrow y_t = \beta_0 + \beta_1 t + \rho e_{t-1} + \epsilon_t \]

\[ y_t = \beta_0 + \beta_1 t + e_t \Rightarrow y_{t-1} = \beta_0 + \beta_1(t - 1) + e_{t-1} \Rightarrow \rho y_{t-1} = \rho \beta_0 + \rho \beta_1(t - 1) + \rho e_{t-1} \]

\[ \Rightarrow y_t - \rho y_{t-1} = \beta_0 + \beta_1 t + \rho e_{t-1} + \epsilon_t - (\rho \beta_0 + \rho \beta_1(t - 1) + \rho e_{t-1}) \]

\[ = \beta_0(1 - \rho) + \beta_1[t - \rho(t - 1)] + \rho e_{t-1} + \epsilon_t - \rho e_{t-1} \]

\[ = \beta_0(1 - \rho) + \beta_1[t - \rho(t - 1)] + \epsilon_t \]
common (McDowall et al., 1980): the dependent variable under study has been operationalized as a time series; a discrete intervention that divides the time series into two distinct segments, one consisting of preintervention observations and one consisting of postintervention observations.

SSD are essentially based on interrupted time series designs (Swaminathan et al., 2014; McDowall et al., 1980). The basic A-B design is made of a continuous time series and a treatment intervention applied in the middle of the series. At the time point when the intervention is applied, there is the change-point of the time series. The change-point divides the continuous time series into a preintervention phase and a postintervention phase. There are two advantages of applying time series analysis in SSD (Kazdin, 1984): (a) the analysis considers autocorrelation and therefore, provides the appropriate test result when the data has serial dependency; (b) the analysis provides interpretations of behavior change across phases.

Analysis of the time series quasi-experiment is a statistical comparison of the pre- and post-intervention time series segments (McDowall et al., 1980). Within each segment of the time series, analyses of stability and trend of the time series provide the foundation of the between phases comparison. Houle (2009) defined the consistently change over time series as a trend. Gorman and Allison (1997) described different types of trends over time: positive trend (systematic increase series); flat series lacking a trend; negative trend (systematic decrease series). On the contrary, the series that return to previous value is a drift (McDowall et al., 1980). Figure 6 demonstrates the four different types of time series patterns.

INSERT FIGURE 6 ABOUT HERE
In interrupted time series data, there are three types of changes between two adjacent time series segments: level change, slope change, and a combination of both (Campbell & Herzinger, 2010). In an A-B design, possible intervention effect patterns are: (a) change in level, no change in slope; (b) no change in level, change in slope; (c) change in level and slope; (d) no change in level or slope. In time series analysis, the concept of slope is discussed as a “trend”. Figure 7 demonstrates the four patterns of intervention effect in SSR.

**INSERT FIGURE 7 ABOUT HERE**

Bayesian Estimation and SSR

Bayesian estimation methods have been applied in SSR data analyses. Zucker et al. (1997) applied a hierarchical Bayesian random effect model in their study of medicine. de Vries and Morey (2013) introduced a new set of Bayes factor tests for single subject data with two phases. Their model took serial dependency into account. Swaminathan et al. (2014) applied Bayesian method in a multilevel model to analyze the A-B-A-B design with multiple subjects, while taking into account the autocorrelation of error terms. Compared with the traditional statistical significance test (SST), Bayesian probability theory has the following advantages:

(a) Bayesian framework expands the single valued hypothesis into a prior distribution (Kruschke, 2011; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Instead of assuming that the parameter of interest has a single value, Bayesian estimation methods allow the probable values of the parameter to be a distribution, i.e. the prior belief \( P_{prior} \sim f(\theta) \). This helps us to keep an open mind of all the possible values of the unknown parameter.
(b) When interpreting the output, the SST inferential is based on the significance level of $p(Data|\theta)$, which is the probability of obtaining the sample data given the null hypothesis is true (Thompson, 1994). In Bayesian framework, $p(Data|\theta)$ is called the likelihood (Kruschke, 2011). Bayesian theory applies the likelihood on the prior distribution of the parameter and updates the prior belief into the posterior distribution [$p(\theta|Data)$, the probability of the model and parameter, based on the given sample data]:

$$p(\theta|Data) \propto p(Data|\theta) \times prior.$$  

(c) When dealing with smaller sample sizes, Bayesian methods are more effective than other estimators (Martin, 2005). Bayesian methods maximize the utility of the sample data by updating the prior distribution through an iteration process, until the parameter estimation stay stable (Kruschke, 2011).

Based on the Bayes factors introduced by de Vries and Morey (2013), the effect size measurement and the level and/or trend effect evaluation is conducted in Bayesian framework. In Bayesian estimation, the prior distribution plays an important role on the parameter estimation and Bayes factors evaluation (de Vries & Morey, 2013; Kruschke, 2011). de Vries and Morey (2013) demonstrated the Cauchy distribution as a prior for $\delta$ (the intervention effect of SSD) and the Beta($a = 1$, $b = 5$) distribution for the autocorrelation ($\rho$).

Bayes Factor in Model Comparison

In Bayesian data analysis, Bayes factor (BF) is used for model comparison (Kruschke, 2011). To compare the applicability of two models ($M1$ and $M2$) on one data
set \((D)\), we investigate the ratio of the posterior probabilities for the two models and apply the Bayes’ rule on both the numerator and denominator (Kruschke, 2011):

\[
p(M_1|D) = \frac{p(D|M_1)p(M_1)}{p(D)} = \frac{p(D|M_1)p(M_1)}{p(D|M_2)p(M_2)} = \frac{p(D|M_1)}{p(D|M_2)} \times \frac{p(M_1)}{p(M_2)}
\]

\(p(M_i|D)\) is the posterior probability for model \(i\) \((i=1, 2)\); \(p(M_i)\) is the likelihood for each model; \(p(M_i)\) is the prior, or the assumption for each model; and \(p(D)\) is the evidence \([p(D) = p(D|M_1)p(M_1) + p(D|M_2)p(M_2)]\). The ratio of \(p(D|M_i)\) is defined as the Bayes factor:

\[
BF = \frac{p(D|M_1)}{p(D|M_2)}
\]

Therefore, Bayes factor is the relative evidences of the sample data under the two models. The evidence for each model is (Wagenmakers et al., 2010):

\[
p(D|M_i) = \int p(D|\theta,M_i)p(\theta|M_i)d\theta
\]

In single subject research designs, assuming that the effect size (intervention effect) of the research is \(\delta\), we compare the null hypothesis model \((H_0: \delta = 0)\) with the alternative hypothesis model \([H_1: \delta \sim p(\delta)]\) by calculating the Bayes factor:

\[
BF = \frac{\int p(D|\theta,H_0)p(\theta|H_0)d\theta d\delta}{\int p(D|\theta,H_1)p(\theta|H_1)d\theta d\delta} = \frac{\int p(D|\theta,\delta = 0)p(\theta)d\theta}{\int p(D|\theta,\delta)p(\theta)p(\delta)d\theta d\delta}
\]

de Vries and Morey (2013) applied Savage-Dickey identity onto the above equation for the estimation of Bayes factor and transferred the ratio of the evidences for different models into the ratio of the marginal posterior distribution to the marginal prior distribution when \(\delta\) equals zero (both under the alternative hypothesis).

\[
BF = \frac{p(D|H_0)}{p(D|H_1)} = \frac{p(\delta = 0|D)}{p(\delta = 0)}
\]
$p(\delta = 0|D)$ is the marginal posterior distribution within the alternative hypothesis ($H_1$); and $p(\delta = 0)$ is the marginal prior distribution within the alternative hypothesis ($H_1$). Morey, Rouder, Pratte and Speckman (2011) introduced a method that used MCMC chain outputs to estimate the Bayes factor. The detailed techniques are out of the range of this review. Interested readers may refer to the original study.

In SSR data analysis, de Vries and Morey (2013) proposed a set of Bayes factors, including Bayes factor for level comparison ($B_{\text{JZS}}$); Bayes factor for level comparison with autocorrelation in the model ($B_{\text{JZS+AR}}$); and Bayes factors for level and trend comparison with autocorrelation ($B_{\text{trend}}$, $B_{\text{int}}$, and $B_{\text{t+t}}$). The Bayes factors can be applied onto the A-B designs, to evaluate the level change under assumption of a flat series ($B_{\text{JZS+AR}}$); and the level and/or trend changes under assumptions of level change only ($B_{\text{int}}$), trend change only ($B_{\text{trend}}$), and both level and trend changes ($B_{\text{t+t}}$).

Summary

Upon reviewing all these features and recent advances of SSD data analysis methods, the conclusion is that using both visual and statistical analysis is the best solution (Houle, 2009). Nowadays, both visual and statistical analysis methods are utilized in SSR. However, visual analysis is still the main analysis method used by most single subject researchers (Gast, 2010; Barlow, Nock, & Herson, 2009). Recent development of statistical analysis methods contributed greatly in the statistical analysis in SSR. However, these methods have not been effectively applied into the practice of single subject researchers and are not very well combined with visual analysis in the research practice of SSD.
The reason is that none of these technical developments reflect a variety of research designs in SSR. In SSR practice, single subject researchers emphasize prominently on the details of the design so that the intervention effect can be verified by the replication (Gast, 2010). It is still in question that how the quantification of intervention effect be evaluated under a time series framework for different types of SSR designs. The current study proposes a series of statistical models using Bayesian estimation to address this problem.
References


Figure 1. *Demonstration for the A-B design*
Figure 2. *Demonstration for the A-B-A-B design*
Figure 3. Demonstration for the multiple baseline design
Figure 4. Demonstration for the alternating treatment design

Figure 5. Model demonstration for the changing criterion design
Figure 6. *Demonstrations for the patterns of time series data*
Figure 7. Demonstrations for the patterns of intervention effect in SSR
GLOSSARY OF TERMS

Autocorrelation - Refers to the correlation between the current data points and the previous data points (i.e., lags).

Convergence - The process of the iteration sequences approaching a limit.

Change-point - The time point when the new phase starts.

HDI (highest density interval) - Refers to an interval that spans most of the distribution, say 95% of it, such that every point inside the interval has higher believability than any point outside the interval.

Level (int) - The intercept of the time series on the y axis or on the vertical line of y = change-point.

Posterior - Computed after taking into account a particular set of observations.

Prior - The researchers' belief before taking into account the sample data.

ROPE (region of practical equivalence) - a small range of values that are equivalent to the value of interest for all practical purposes.

Run length - Number of iteration.

Stability - The standard deviation of the time series.

Trend - The slope of the time series.


