Single subject or n-of-1 research designs have been widely used to evaluate treatment interventions. Many statistical procedures such as split-middle trend lines, regression trend line, Shewart-chart trend line, binomial tests, randomization tests and Tryon C-statistics have been used to analyze single-subject data, but they fail to control Type I error due to serially-dependent time-series observations. The interrupted time series analysis maintains Type I error but assumes that the intervention effect to be a linear trend change from baseline. In this paper, we consider an improved intervention analysis model (Box and Tiao, 1975) for dynamic characteristics of an intervention effect in a short series of single-subject data. The maximum likelihood estimates are derived and a hypothesis testing procedure is proposed. The method is illustrated with a real clinical trial on constraint induced language therapy for aphasia patients.

Keywords: intervention analysis; exact likelihood; single-subject

1. Introduction

Single-subject designs have been used widely for decades, particularly in the behavioral sciences, but statistical analysis of such studies remains problematic, primarily because such data are generally autocorrelated and the observation series is short [1, 2]. The purpose of this paper is to introduce a new analysis model that improves the current methods of drawing conclusions from single-subject studies. In a typical single-subject design, repeated observations are made on the lone subject during a baseline period and a subsequent treatment period. The baseline measurements are intended to establish a stable reference point, and also, in cases of recent injury or other affliction from which spontaneous improvement might be expected, to estimate the rate of improvement prior to treatment. After the subject is exposed to the intervention, the observations continue in an attempt to establish corresponding treatment-period values. Investigators then desire to test two null hypotheses: 1) there is no difference in overall outcome between the baseline and treatment periods, and 2) there is no difference in the rate of change in outcome between the two periods. When spontaneous improvement before treatment does not occur (i.e., the slope of the baseline data is non-positive), rejection of the first hypothesis is enough to show that the treatment is effective. In other cases, rejection of both hypotheses is required [1].

Visual analysis is the traditional and still widely used method of approaching such studies [1]. The data are plotted across time, with a vertical line separating the baseline and treatment periods. Investigators then eyeball the data and make informal conclusions about the effectiveness of the intervention. As one might imagine, this method is highly subjective and hence unreliable, with one large meta-analysis finding an overall inter-rater agreement coefficient of only 0.58 [3]. To address this concern, researchers have proposed various tools to aid visual analysis and make its conclusions more robust. In the split-middle trend line method [4], the baseline data are divided in half and a line is drawn through their respective medians. The same procedure is applied to the treatment data, and the level and slope of the two lines are qualitatively compared. The celeration trend line method [5] is identical, except the lines extend through means rather than medians. Separate regression lines through the baseline and treatment data also are sometimes plotted and visually compared. None
of these methods, however, has been shown to offer much improvement in the reliability of visual analysis, with Type I error rates remaining as high as 84% [6, 7, 8, 9, 10].

Methods that are more statistically oriented also often are used. The Shewart procedure [5, 11] sets reference lines two standard deviations above and below the mean of the baseline data. If two successive data points in the treatment period fall outside those bounds, one infer that significant change has occurred. Binomial tests compare the proportion of treatment points that fall above and below the baseline split-median or celeration line. T-tests are sometimes performed between the baseline and treatment means, and Tryon’s C-statistic [12] is frequently used to compare slopes. The autocorrelation of single-subject data causes such tests to be invalid, however, and Type I error rates remain unacceptably high when autocorrelation is present [13, 14, 15, 16].

Gottman [17] defined interrupted-time-series analysis (ITSA) for a stream of serially dependent observations across two experimental periods. Based on fitting autoregressive parameters, the method yields three tests: an F test of the null hypothesis that no overall change has occurred between the two periods, and t-tests for differences in means and slopes. Crosbie [18] showed that the ITSA method underestimated positive autocorrelation and hence could not maintain Type I error control when the baseline and treatment observations were relatively few (less than 50 observations per time period), making the ITSA method inapplicable to most clinical settings. Crosbie proposed a corrected version of ITSA, called ITSACORR, which could handle shorter time series and has since been widely employed. ITSACORR, however, fails to control Type I error for autocorrelations higher than 0.6 and sample sizes less than 20, and it also assumes that the intervention effect is a linear trend change from baseline, which is not appropriate in many applications, such as the examples we give in our illustrative example section. In addition, Rosner, Munoz, et al. [19] and Rosner and Munoz [20] present autoregressive models that use regression methods to relate change in response variables to explanatory variables.

In this paper, we consider an improved intervention analysis model [21] for dynamic characteristics of an intervention effect in a short series of single-subject data. The statistical model is presented in Section 2. The maximum likelihood estimates are derived and a hypothesis testing procedure is proposed in Section 3. The methods are illustrated with a real clinical trial on constraint-induced language therapy for aphasia patients in Section 4. In section 5 we presented some concluding remarks.

2. The intervention analysis model

Box and Tiao [21] considered, among others, the following intervention model

\[
Y_t = f(t) + N_t, \tag{1}
\]

where \(Y_t\) is the observed outcome series, \(f(t)\) is the unknown mean function that follows a first-order dynamic model for intervention, and \(N_t\) is random noise that follows a mixed autoregressive moving average (ARMA) process. Specifically, \(f(t)\) is assumed to be a transfer function of the form

\[
f(t) = \frac{\omega B}{1 - \delta B} I_t,
\]

where \(\omega\) and \(\delta\) are unknown parameters for intervention effect, \(B\) is the backward shift operator [22] and an indicator function \(I_t\) is given by

\[
I_t = \begin{cases} 0, & \text{if } t < T; \\ 1, & \text{o/w}. \end{cases}
\]

This implies that \(f(t) = \omega + \delta f(t - 1)\) when \(t\) is after intervention starting time \(T\); in other words,

\[
f(t) = \begin{cases} 0, & \text{if } t < T; \\ \omega(1 - \delta^{t-T})/(1 - \delta), & \text{if } t > T, \end{cases}
\]

with steady state gain of \(\omega/(1 - \delta)\). In this intervention analysis model the first order dynamic function is applied to the unknown mean function, which makes it hard to derive maximum likelihood estimate because the parameters are involved “non-linearly” in the model. Usually, least squares values are used, and they are close approximate to the MLE when the time series is long enough, but this is usually not the case in single subject analysis.
In this paper, we consider applying first order dynamic directly to the observed time series. Specifically, we propose the following model:

\[
\begin{align*}
Y_1 &= \mu + W_1 \\
Y_2 &= \mu + W_2 \\
&\vdots \\
Y_T &= \mu + W_T \\
Y_{T+1} &= \mu + \omega + \delta Y_T + W_{T+1} \\
&\vdots \\
Y_n &= \mu + \omega + \delta Y_{n-1} + W_n,
\end{align*}
\]  

(3)

where \(W_t\) follows an ARMA(1, 1) model with mean zero. Note that higher order \(W_t\) can be included into the model, but our experiences suggest that ARMA(1, 1) seems sufficient for short time series data in single subject analysis. Specifically, we assume that

\[
W_t = \frac{1 - \theta B}{1 - \phi B} \alpha_t
\]  

(4)

which implies that \(W_t - \phi W_{t-1} = \alpha_t - \theta \alpha_{t-1}\).

**Remark 1.** Under model (1), we have \(Y_t = f(t) + N_t = \omega + \delta f(t-1) + N_t = \omega + \delta Y_{t-1} + (N_t - \delta N_{t-1})\) when \(t \geq T + 1\). Therefore, if we let \(W_t = N_t - \delta N_{t-1} - E(N_t - \delta N_{t-1})\) for \(t \geq T + 1\) and \(W_t = N_t - E(N_t)\) for \(t \leq T\), then the original intervention model (1) can be written in the form of (3). However, one can see that in this new form, the shape parameter \(\delta\) appears in both the intervention effect and the error term.

The above model (3) can be rewritten in matrix form as

\[
W = Y - B\beta = AY - \eta,
\]  

(5)

where \(\beta = (\mu, \omega, \delta)'\), \(\eta = (\mu, \ldots, \mu, \omega, \ldots, \mu + \omega)'\), and matrices \(A\) and \(B\) given as follows

\[
A = \begin{bmatrix}
1 & 0 & 0 & \ldots & \ldots & 0 & 0 \\
0 & 1 & 0 & \ldots & \ldots & 0 & 0 \\
& \vdots & 0 & \ldots & \ldots & -\delta & 1 \\
& 0 & 0 & \ldots & \ldots & -\delta & 1 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
\vdots & \vdots & \vdots \\
1 & 1 & Y_T \\
\vdots & \vdots & \vdots \\
1 & 1 & Y_{n-1}
\end{bmatrix}.
\]  

(6)

As we will see in the next section, the new model is considerably simpler and allows us to derive exact MLEs when we only have limited few observations.

### 3. The estimation procedure

It has been shown by Newbold [23], that density of \(W\) is given by

\[
p(W|\phi, \theta, \sigma) = (2\pi\sigma^2)^{-\frac{p}{2}}|Z'Z|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}S(\phi, \theta)/\sigma^2\right\}
\]  

(7)

where \(Z\) is given by

\[
Z = \begin{bmatrix}
1 & 0 & 0 \\
0 & \theta - \phi & 1 \\
\theta(\theta - \phi) & -\phi(\theta - \phi)(1 - \phi^2)^{-\frac{1}{2}} & 1 \\
\theta^2(\theta - \phi) & -\theta^2\phi(\theta - \phi)(1 - \phi^2)^{-\frac{1}{2}} & \theta(\theta - \phi) \\
\vdots & \vdots & \vdots \\
\theta^{n-1}(\theta - \phi) & -\theta^{n-1}\phi(\theta - \phi)(1 - \phi^2)^{-\frac{1}{2}} & \theta^{n-1}(\theta - \phi) \\
\theta^n(\theta - \phi) & -\theta^n\phi(\theta - \phi)(1 - \phi^2)^{-\frac{1}{2}} & \theta^n(\theta - \phi)
\end{bmatrix}.
\]  

(8)
and \( S(\phi, \theta) = W'L'(I - Z(Z'Z)^{-1}Z')LW \) with matrix \( L \) defined as

\[
L = \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 \\
1 & 0 & 0 & \ldots & 0 & 0 \\
(\theta - \phi) & 1 & 0 & \ldots & 0 & 0 \\
\theta - \phi & (\theta - \phi) & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\theta^n - 2(\theta - \phi) & \theta^n - 3(\theta - \phi) & \theta^n - 4(\theta - \phi) & \ldots & (\theta - \phi) & 1
\end{bmatrix}.
\] (9)

Clearly, that the value of matrices \( Z \) and \( L \) depend only on the parameters of the time series \( W_t \).

Since the Jacobian of the transformation \( W = AY - \eta \) is \(|\det(A)| = 1\), the probability density function for the observed data \( Y = (Y_1, \ldots, Y_n) \) is given as follows

\[
p(Y|\phi, \theta, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}}|Z'Z|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} S^*(\phi, \theta)/\sigma^2 \right\},
\] (10)

where \( S^*(\phi, \theta, \beta) = (AY - \eta)'\Gamma(AY - \eta) = (Y - B\hat{\beta})'\Gamma(Y - B\hat{\beta}) \) and \( \Gamma = L'(I - Z(Z'Z)^{-1}Z')L \).

Now consider the likelihood function:

\[
l(\phi, \theta, \beta, \sigma^2|Y) = (2\pi\sigma^2)^{-\frac{n}{2}}|Z'Z|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} S^*(\phi, \theta, \beta)/\sigma^2 \right\}.
\] (11)

First, it can be shown that for any given \((\phi, \theta)\), the likelihood function is maximized by \( \hat{\beta}(\phi, \theta) = (B'\Gamma B)^{-1}B'\Gamma Y \) and \( \hat{\sigma}^2(\phi, \theta) = S^*(\phi, \theta, \hat{\beta})/n \), which depend on \((\phi, \theta)\) through \( \Gamma \). Plugging them into the likelihood function, we get

\[
l^*(\phi, \theta, \hat{\beta}, \hat{\sigma}^2, Y) = \left( \frac{2\pi \cdot S^*(\phi, \theta, \hat{\beta})}{n} \right)^{-\frac{n}{2}} |Z'Z|^{-\frac{1}{2}} \exp \left\{ -\frac{n}{2} \right\}.
\] (12)

Therefore, if we let \((\hat{\phi}, \hat{\theta})\) be the values that maximize the above expression \( l^* \), then the MLE of the parameters can be obtained as \( \hat{\phi}, \hat{\theta}, \hat{\beta}(\hat{\phi}, \hat{\theta}) \) and \( \hat{\sigma}^2(\hat{\phi}, \hat{\theta}) \).

Furthermore, we would like to point out a connection between the MLE \( \hat{\beta}(\hat{\phi}, \hat{\theta}) \) and a Bayes estimator. If we let \( \hat{\beta} = (B'\Gamma B)^{-1}B'\Gamma Y \) which depend on \((\phi, \theta)\) through \( \Gamma \), then we have

\[
S^*(\phi, \theta, \beta) = [(Y - B\hat{\beta}) - B(\beta - \hat{\beta})]'\Gamma[(Y - B\hat{\beta}) - B(\beta - \hat{\beta})]
= (Y - B\hat{\beta})'\Gamma(Y - B\hat{\beta}) + (\beta - \hat{\beta})'B'\Gamma B(\beta - \hat{\beta}),
\] (13)

where the first term is constant given \((\phi, \theta)\) and \( Y \). Therefore, the likelihood function and (13) imply that, conditioned on \((\phi, \theta, \sigma^2)\), the posterior distribution of \( \beta \) is multivariate normal with mean \( \hat{\beta} \) and covariance \((B'\Gamma B)^{-1}\). In other words,

\[
p(\beta|Y, \phi, \theta, \sigma^2) \propto \exp \left\{ -\frac{1}{2} (\beta - \hat{\beta})'B'\Gamma B(\beta - \hat{\beta})/\sigma^2 \right\}.
\] (14)

In other words, the MLE \( \hat{\beta}(\hat{\phi}, \hat{\theta}) \) is the Bayes estimator with auxiliary parameters estimated at \((\hat{\phi}, \hat{\theta})\).

Suppose \( \beta' = (\beta_1', \beta_2') \), where \( \beta_2 = (\omega, \delta) \) is the treatment effect of interest. In other words, we would like to test the following hypothesis

\[
H_0 : \beta_2 = (0, 0) \quad \text{vs} \quad H_a : \beta_2 \neq (0, 0)
\] (15)

It is well known, that maximum likelihood estimate \( \hat{\beta}_2 \) has asymptotic normal distribution, with mean equals to the true value of \( \beta_2 \) and variance-covariance matrix can be consistently estimated by the lower 2 by 2 matrix of \((B'\Gamma B)^{-1}\). Therefore, when there is large sample size \( n \), one may employ a Wald test, which is based on \( T_w = \hat{\beta}_2'\Sigma \hat{\beta}_2 \) that has chi-squared distribution with 2 degrees of freedom.

Since this result may not be valid for small samples, we propose the following procedure to find the p-value:

1. Simulate \( n \) time series from the model with intervention effect \( \beta_2 = (0, 0) \) and noise parameters \((\hat{\phi}, \hat{\theta})\).
2. Estimate coefficients \( \beta_2 \) based on simulated data and calculate \( T_i, i = 1, \ldots, n \) as a Wald test statistic.
3. Estimate \( p \)-value according the formula

\[
\hat{p} = \frac{\sum_{i=1}^{n} I(T_i > T_w)}{n}.
\]
4. An illustrative example and a simulation study

To show how the model from Section 2 can be applied to short time series with an intervention, we consider data from a randomized clinical trial of Constraint Induced Language Therapy (CILT). The main aim of the study was to determine if CILT would result in observable improvements in speech and if it would be significantly better than regular, unconstrained language therapy. There were 36 patients who completed the study: ten participants in each of the two CILT groups and eight participants in each of the two Promoting Aphasic Communicative Effectiveness (PACE) groups. We applied the model from Section 2, with the assumption that errors of the true model are from $AR(1)$ process and the time of the intervention is known, for each patient in each group separately. The assumption of $AR(1)$ process was chosen because there was no big improvement with $ARMA(1,1)$ model. Results of all patients from the clinical trial are summarized in Table 1.

Figure 1 shows the result of fitting the proposed model to the Intensive PACE group of patients. Parameter estimates and p-values from permutation test for the 8 patients from the Intensive PACE group are presented in the top 8 rows of the Table 1. From the plots in Figure 1, we observed high variability in the responses on the intervention within the group. Such variability was observed in other three groups of patients as well.

In addition, a simulation study was conducted to compare models with $AR(1)$ and $ARMA(1,1)$ errors. The performance of the models was tested based on 1000 time series of 13 observations, each of them generated from a model with parameters $\beta = (30, 8, 0.6)$ and known time of intervention. For the comparison purpose we looked at the two sets of data. The first set was simulated using the model with $AR(1)$ errors and the second set was simulated using the model with $ARMA(1,1)$ errors. We applied the model from Section 2 with assumption that errors follow $AR(1)$ and $ARMA(1,1)$ processes to each series in both data sets. In this simulations study, we consistently observed that the model with the assumption that errors follow $AR(1)$ process slightly outperformed the model with the assumption of $ARMA(1,1)$ process in terms of mean prediction error (MPE). In Table 3 we present results of the simulations, which were performed in R 2.15.1 [24].

5. Discussion

In this paper we have presented a modeling framework based on an improved intervention analysis model for single-subject studies with relatively small number of observations for each subject. The small number of observation is a potential difficulty in single-subject studies. To address this issue we simplified the intervention analysis model proposed by Box and Tiao [21] and derived exact likelihood function, which allows us to get estimates in more direct way than approximate algorithms, which often fail to converge for short series.

One potential limitation is that two different sets of parameters $(\omega, \delta)$ can describe a similar change pattern in the mean function. The simplest way to address the limitation would be to impose constraints on parameters $(\omega, \delta)$. However, our
Table 1. Case study results

| Patient | p-value | \( \mu \)  | \( \omega \)  | \( \delta \)  |
|---------|---------|------------|-------------|------------|
| 1       | < 0.001 | 56.177     | 1.31        | 0.325      |
| 2       | 0.87    | 42.51      | -12.845     | 0.25       |
| 3       | < 0.001 | 9.25       | 34.82       | 0.21       |
| 4       | 0.001   | 60.41      | 17.56       | 0.06       |
| 5       | < 0.001 | 20.02      | 18.98       | 0.47       |
| 6       | 0.041   | 56.84      | 51.33       | -0.48      |
| 7       | 0.689   | 3.98       | 2.75        | 0.047      |
| 8       | 0.014   | 47.36      | 1.14        | 0.21       |
| 9       | < 0.001 | 19.5       | -0.12       | 0.69       |
| 10      | 0.002   | 59.06      | -28.76      | 0.62       |
| 11      | 0.23    | 64.18      | 44.66       | -0.495     |
| 12      | 0.003   | 13.15      | 0.79        | 0.65       |
| 13      | < 0.001 | 50.07      | 1.40        | 0.35       |
| 14      | < 0.001 | 46.52      | 13.61       | 0.33       |
| 15      | 0.035   | 56.13      | -34.16      | 0.7        |
| 16      | < 0.001 | 34.10      | -18.72      | 0.72       |
| 17      | 0.01    | 3.58       | 4.39        | 0.56       |
| 18      | < 0.001 | 37.28      | -16.7       | 0.76       |
| 19      | < 0.001 | 9.41       | 39.59       | 0.13       |
| 20      | 0.015   | 69.4       | -29.98      | 0.52       |
| 21      | < 0.001 | 10.41      | 23.58       | 0.44       |
| 22      | < 0.001 | 33.17      | 20.57       | 0.38       |
| 23      | 0.031   | 34.44      | -7.68       | 0.56       |
| 24      | < 0.001 | 53.51      | -20.22      | 0.62       |
| 25      | < 0.001 | 2.01       | 41.61       | -0.03      |
| 26      | < 0.001 | 7.83       | 49.02       | 0.09       |
| 27      | 0.37    | -2.41      | 6.41        | -0.22      |
| 28      | 0.001   | 17.86      | 5.5         | 0.16       |
| 29      | 0.026   | 3.97       | -2.75       | 0.25       |
| 30      | 0.089   | 4.34       | -0.94       | 0.55       |
| 31      | 0.015   | 23.10      | -17.89      | 0.44       |
| 32      | 0.011   | 16.92      | 28.99       | 0.19       |
| 33      | 0.354   | -0.02      | 0.79        | -0.75      |
| 34      | 0.64    | 0.17       | -0.034      | 0.37       |
| 35      | 0.024   | 39.68      | 44.85       | -0.33      |
| 36      | < 0.001 | 31.26      | 16.64       | 0.38       |

Table 2. Mean Prediction Error and Power for models with \( \beta = (30, 8, 0.6) \) and errors from AR(1) and ARMA(1,1) processes

| Model         | Fit  | MPE  | Power |
|---------------|------|------|-------|
| AR(1)         | AR(1)| 1.28 | 1     |
| ARMA(1,1)     | AR(1)| 1.49 | 1     |
| ARMA(1,1)     | AR(1)| 1.94 | 1     |
| ARMA(1,1)     | ARMA(1,1)| 2.02 | 1     |

Simulation results confirmed that this limitation does not affect the testing procedure regarding a null hypothesis of no intervention effect.

There are several possible extensions of the proposed model. First and immediate, when instead of one intervention the model can be applied to the clinical studies with multiple interventions during a study period. The ease of implementation
Table 3. Mean Prediction Error and Power for models with $\beta = (0, 0, 0)$ and errors from AR(1) and ARMA(1,1) processes

| Model       | Fit   | MPE  | Type I Error |
|-------------|-------|------|--------------|
| AR(1)       | AR(1) | 1.087| 0            |
| ARMA(1,1)   | AR(1) | 1.22 | 0            |
| ARMA(1,1)   | AR(1) | 1.19 | 0            |
|             | ARMA(1,1) | 1.334 | 0         |

will encourage adoption of this methodology in medical research. The model fitting has been implemented both in R and in MATLAB and the program codes are available upon request. As an illustrative example, the model was successfully fit to the data from a randomized clinical trial of Constraint Induced Language Therapy. Lastly, the applications of such models are not restricted to clinical studies.

References

1. Robey R, Schultz M, Crawford A, Sinner C. Review: Single-subject clinical-outcome research: designs, data, effect sizes, and analyses. *Aphasiology* 1999; 13(6):445–473.
2. Lillie EO, Patay P, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med.* 2011; 8(2):161–173.
3. Ottenbacher K. Interrater agreement of visual analysis in single-subject decisions: Quantitative review and analysis. *American Journal on Mental Retardation* 1993.
4. Kazdin A. *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press New York, 1982.
5. Bloom M, Fischer J, Orme J, et al. *Evaluating Practice*. Allyn & Bacon, 1982.
6. Matyas T, Greenwood K. Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis* 1990; 23(3):341.
7. Johnson M, Ottenbacher K. Trend line influence on visual analysis of single-subject data in rehabilitation research. *Disability & Rehabilitation* 1991; 13(2):55–59.
8. Ottenbacher K, Cusick A. An empirical investigation of interrater agreement for single-subject data using graphs with and without trend lines. *Journal of the Association for Persons with Severe Handicaps* 1991.
9. Stocks J, Williams M. Evaluation of single subject data using statistical hypothesis tests versus visual inspection of charts with and without celeration lines. *Journal of Social Service Research* 1995; 20(3-4):105–126.
10. Ottenbacher K. Visual inspection of single-subject data: An empirical analysis. *Mental Retardation* 1990.
11. Krishef C. *Fundamental approaches to single subject design and analysis*. Krieger Pub. Co., 1991.
12. Tryon W. A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis* 1982; 15(3):423.
13. Phillips J. Serially correlated errors in some single-subject designs. *British Journal of Mathematical and Statistical Psychology* 1983; 36(2):269–280.
14. Tootaher L, Banz M, Noble C, Camp J, Davis D. N=1 designs: The failure of anova-based tests. *Journal of Educational and Behavioral Statistics* 1983; 8(4):289–309.
15. Sharpley C, Alavossis M. Autocorrelation in behavioral data: An alternative perspective. *Behavior Assessment* 1988; 10:243–251.
16. Suen H, Lee P, Owen S. Effects of autocorrelation on single-subject single-facet crossed-design generalizability assessment. *Behavioral Assessment* 1990; 12:305–315.
17. Gottman J. *Time-series analysis: A comprehensive introduction for social scientists*, vol. 400. Cambridge University Press Cambridge, 1981.
18. Crosbie J. Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology* 1993; 61(6):966.
19. Rosner B, Muñoz A, Tager I, Speizer F, Weiss S. The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. *Statistics in Medicine* 2006; 4(4):457–467.
20. Rosner B, Muñoz A. Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations. *Statistics in Medicine* 1988; 7(1-2):59–71.
21. Box G, Tiao G. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* 1975; 70:70–79.
22. Box G, Jenkins G, Reinsel G. *Time series analysis: forecasting and control*. Prentice Hall, 1994.
23. Newbold P. The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika* 1974; 61(3):423–426.
24. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2013. URL http://www.R-project.org/.