Journal of Modern Applied Statistical Methods

Volume 6 | Issue 2 Article 4

11-1-2007

# Time-Series Intervention Analysis Using ITSACORR: Fatal Flaws

Bradley E. Huitema
*Western Michigan University*

Joseph W. McKean
*Western Michigan University*

Sean Laraway
*San Jose State University*

Journal of Modern Applied Statistical Methods
November, 2007, Vol. 6, No. 2, 367- 379

1538 – 9472/05/$95.00

# Time-Series Intervention Analysis Using ITSACORR: Fatal Flaws



Bradley E. Huitema Joseph W. McKean
Western Michigan University

Sean Laraway
San Jose State University

The ITSACORR method (Crosbie, 1993, 1995) is evaluated for the analysis of two-phase interrupted time-series designs. It is shown that each component of the ITSACORR framework (including the structural model, the design matrix, the autocorrelation estimator, the ultimate parameter estimation scheme, and the inferential method) contains fatal flaws.

Key words: Autocorrelation, time-series intervention analysis, time-series regression with auoregressive errors.

## Introduction

Researchers and practitioners working in the behavioral sciences frequently employ interrupted time-series designs to determine the effectiveness of various interventions in both clinical and natural settings. Currently, several methods are available for statistically analyzing data from interrupted time-series designs. Among these methods, autoregressive integrated moving average (ARIMA) intervention models have a long history of endorsement by methodologists (e.g., Glass, Willson, & Gottman, 1975; McCleary & Hay, 1980). Nevertheless, some authors (e.g., Gorman & Allison, 1997) have noted that certain properties of ARIMA models, particularly their analytical complexity and requirement of relatively large sample sizes, make the use of these models troublesome for many behavioral researchers. Concerns regarding these undesirable properties of ARIMA models have prompted the development of several alternatives. These alternatives reportedly (a) reduce the difficulty of analyzing time-series data and (b) enable the analysis of series with relatively few observations, a characteristic of many applications of time-series designs in the

Bradley E. Huitema is Professor of Psychology. His major interests are in the design and analysis of single-case experiments, quasi-experiments, and observational studies. Email: brad.huitema@wmich.edu. Joseph W. McKean is Professor of Statistics. His interests are in robust nonparametric statistical methods. Sean Laraway is Assistant Professor of Psychology. His research interests include behavioral pharmacology and learning and memory.

behavioral sciences. Two commonly cited alternatives to ARIMA intervention models are Gottman's ITSE (Gottman,1981; Rushe & Gottman, 1993) and Crosbie's ITSACORR (Crosbie, 1993, 1995).

Both of these alternatives use the same underlying model and estimate the same intervention parameters. Despite recent corrections, the current version of ITSE does not provide a satisfactory method for analyzing time-series data because it still contains several major defects. These defects are not software bugs; rather, they are problems with the method that are described in a recent critique (Huitema, 2004).

The ITSACORR method builds on the ITSE method; it was designed to analyze short series that likely have autocorrelated errors and that may have trend within one phase or within both phases (Crosbie, 1995). In proposing ITSACORR as a suitable method for analyzing time-series data, Crosbie (1993, 1995) described several supposed advantages of ITSACORR over both ARIMA intervention methods and Gottman's ITSE. First, unlike ARIMA, ITSACORR allegedly yields appropriate results with small sample sizes even in the presence of high levels of autocorrelation (Crosbie, 1995, p. 392). Second, ITSACORR reportedly provides results that agree with those of ARIMA when a large number of observations is available (Crosbie, 1995, pp. 391-392). Third, ITSACORR supposedly has better small-sample inferential properties than does ITSE (Crosbie, 1995).

These claims combined with readily available and uncomplicated software have led to considerable attention for ITSACORR from methodologists and practitioners. Writers in applied fields such as aphasiology, applied behavior analysis, clinical psychology, counseling psychology, and school psychology have strongly encouraged its use. For example, Gottman and Rushe (1993) described ITSACORR as "a new, powerful method for single-case analysis of change over time using the interrupted time-series design . . . this can be done without needing to know sophisticated time-series modeling methods and with very few data before and after the intervention" (p. 909). They further state that ITSACORR ". . . makes time-series methods available to the general clinician for the first time" and that "This approach will have widespread importance in the evaluation of change in patients in clinical trials where it is possible to study people on a case-by-case basis, or in the case work of quantitatively oriented clinical practitioners" (p. 909). This initial endorsement has been followed by additional support (e.g., Gottman, 1995), and ITSACORR has received many positive evaluations published in single-case methodology books (e.g., Franklin, Allison, & Gorman, 1997). Gorman and Allison (1997), for instance, have stated that ITSACORR "combines the best of ARIMA and regression approaches" (p. 94). Similarly, a widely used research methodology textbook (Christensen, 2007) states (p. 345) that Crosbie's method is an effective replacement for the well established methods of Box and Jenkins (1970), Box and Tiao (1965), and Glass, Willson, and Gottman (1975).

In addition to these recommendations from methodologists, ITSACORR has received additional endorsement in expository articles written for practitioners. For example, researchers in the area of aphasiology have stated that "ITSACORR should be the procedure-of-choice, and essentially the standard, for applying hypothesis testing logic to single-subject data" (Robey, Schultz, Crawford, & Sinner, 1999, p. 466). Several other authors (some outside the behavioral sciences) have cited ITSACORR as one of several credible methods for time-series analysis (e.g., Ellis, 1999, p. 573; Hogenraad, McKenzie, & Martindale, 1997, pp. 433-35).

A recent expository article on the design and analysis of time-series studies appeared in The International Journal of Clinical and Experimental Hypnosis; it includes the following endorsement: "ITSACORR is eminently easy to use; it corrects for autocorrelation; it generates statistics that are familiar to reviewers and editors; and it is acceptable for use with as few as 7 to 10 data points per phase" (Borckardt & Nash, 2002, p.127). Following this and other statements, the article presents a half-dozen examples of the use of ITSACORR (pp. 132-142).

It appears that the effect of these books and articles has been widespread acceptance of ITSACORR. One can find many published examples of the application of ITSACORR in journals such as *Aphasiology* (e.g., Robey et al. 1999; Spencer, Doyle, McNeil, Wambaugh, Park, & Carroll, 2000), *British Journal of Clinical Psychology* (e.g., Davidson & Tyrer, 1996), *Journal of Consulting and Clinical Psychology* (e.g., Lucyshyn, Albin, & Nixon, 1997), and *School Psychology Review* (e.g., Stage & Quiroz, 1997). Because ITSACORR is widely recommended and used the descriptive and inferential properties of this method must be understood by methodologists, research workers, and journal editors. The purpose of this article is to explicate these properties.

## Logic of the Two Phase Design

An understanding of the essential descriptive properties associated with the analysis of the interrupted time-series experiment rests on the logic of this design. Consider the simple two-phase (A-B) interrupted time-series design. The data of the first phase can provide a prediction of what would occur during the second phase in the absence of an intervention. The researcher's interest lies in the difference between the predicted (counterfactual) second phase behavior and the behavior that actually occurs during the second phase. There exist two major statistics that characterize this difference. The first is known as level change and the second is known as slope change. Although the interpretation of both of these measures is straightforward, level change is frequently misunderstood and incorrectly computed (Huitema & McKean, 2000a; Huitema, 2004).

## Level Change

One possible measure of level change indicates the amount by which the intervention changes the expected value of the response at the beginning of the intervention phase. If there are n1 observations in the first phase and n2 observations in the second phase, the first observation in the intervention phase occurs at time n1 + 1. The level change can reasonably be defined (under the assumption that an adequate model describes the data for each phase) as the difference between (a) the predicted (counterfactual) value of Y at time n1 +1 based on a model of the first phase data and (b) the expected value of Y at time n1 +1 based on a model of the second-phase data. It is crucial to understand that both of these estimates must be associated with exactly the same time point (viz., n1 +1). Although various time-series intervention models may use different procedures to compute the two level estimates, all acceptable procedures estimate level change at a common time point. It is important to be aware that the concept of level change does not, in general, refer to the difference between the means of the two phases. Level change refers to a shift in elevation that is unexplained by possible within-phase trends.

## Slope Change

Slope change provides the second major way of characterizing the effect of an intervention. Here the term slope has its traditional meaning. It simply refers to the average change in Y given a one-unit change in X, where the X variable is time. If the intervention has an effect, it may produce a change in level, a change in slope, or both. Because a reasonable representation of intervention effects often requires measures of both level change and slope change, an adequate descriptive analysis will usually provide accurate estimates of both of them. Although interventions can also interrupt the structure of time-series data by changing the variance or in other more subtle ways (see, e.g., Stoline, Huitema, & Mitchell, 1980), level change and slope change provide two of the most basic effect measures. The adequacy of ITSACORR with respect to these measures is the focus of this article.

# Methodology

Four linked issues that are relevant in evaluating the adequacy of intervention analyses were studied. First, at the most elementary level, whether ITSACORR produces measures that are consistent with the logic of time-series intervention designs was evaluated. Second, the consistency between the logic of the design and the ITSACORR structural model was examined. Third, the consistency between the ITSACORR structural model and the ITSACORR design matrix was evaluated. Last, the inferential properties of the tests provided by ITSACORR was evaluated. Details regarding these issues and methods used to study them are described in this section.

Correspondence Between the Logic of the Design and the Parameter Estimates Produced by ITSACORR

The correspondence of the level change and slope change estimates produced by ITSACORR with level change and slope change estimates produced by methods that are consistent with the logic of the interrupted time-series design was evaluated. Three methods that are known to provide parameter estimates consistent with the logic of the interrupted time-series design utilize the same design matrix. This design matrix differs greatly from the matrix used by both ITSE and ITSACORR. Described is the appropriate matrix (denoted as the H-M matrix) in detail elsewhere (e.g., Huitema & McKean, 2000a, 2000b; Huitema, McKean, & McKnight, 1994; McKnight, McKean, & Huitema, 2000). The three methods that use the H-M matrix differ from each other in terms of assumptions and/or method of estimation. The first method (H-M OLS) assumes independent errors and uses ordinary least-squares (OLS) as its estimation procedure. Although some researchers believe that OLS models are never appropriate in the case of time-series designs, this is not true (see Huitema and McKean, 1998). The second and third methods assume first-order autoregressive errors. They differ from each other in that the second method (H-M M-L) uses a maximum-likelihood estimation procedure, whereas the third method (H-M Bootstrap) uses a double bootstrap approach (McKnight et al., 2000).

After results from the first three methods were obtained ITSACORR was applied to the same data and made comparisons among the results of the different methods. All of these comparisons used data from four published studies (see Figure 1). These data are of the type for which ITSACORR was specifically designed. Indeed, all of these data were obtained from expository articles that illustrate and promote the use of ITSACORR (i.e., Borckardt, 2002; Crosbie, 1995; Robey et al., 1999; Spencer et al., 2000).

Correspondence Between the Logic of the Intervention Design and the ITSACORR Structural Model

The evaluation of how well the ITSACORR model corresponds to the logic of the interrupted time-series design involved comparing the level- and slope-change parameters defined in the structural model with the change parameters of interest in the intervention design. This involved answering two questions: (a) Does the ITSACORR model define level change as the difference between the counterfactual level and the observed level? and (b) Does the model define slope change as the difference between the counterfactual slope and the observed slope?

Correspondence Between the Structural Model and the Design Matrix

A coherent methodology will have consistency between the parameters specified in the structural model and the parameters implied by the associated design matrix. This consistency was evaluated by comparing the level change, slope change, and first order autocorrelation parameters specified in the ITSACORR structural model with the corresponding parameters defined by the ITSACORR design matrix.

Evaluation of Inferential Performance

ITSACORR provides inferential tests on the difference between intercepts and slopes.

The inferential aspects of greatest interest in evaluating the performance of hypothesis testing procedures are Type I error and power. A small computer simulation was used to empirically evaluate these properties. The simulation study evaluated these properties under two levels of autocorrelation (.50 and .80) and two intercept change effect sizes (0 and 10 sigma); total sample size ($n1 + n2$) was set at 20. No slope change was included in any of the simulations. 1,000 simulations were performed under each condition; $a$ was set at the nominal value of .05.

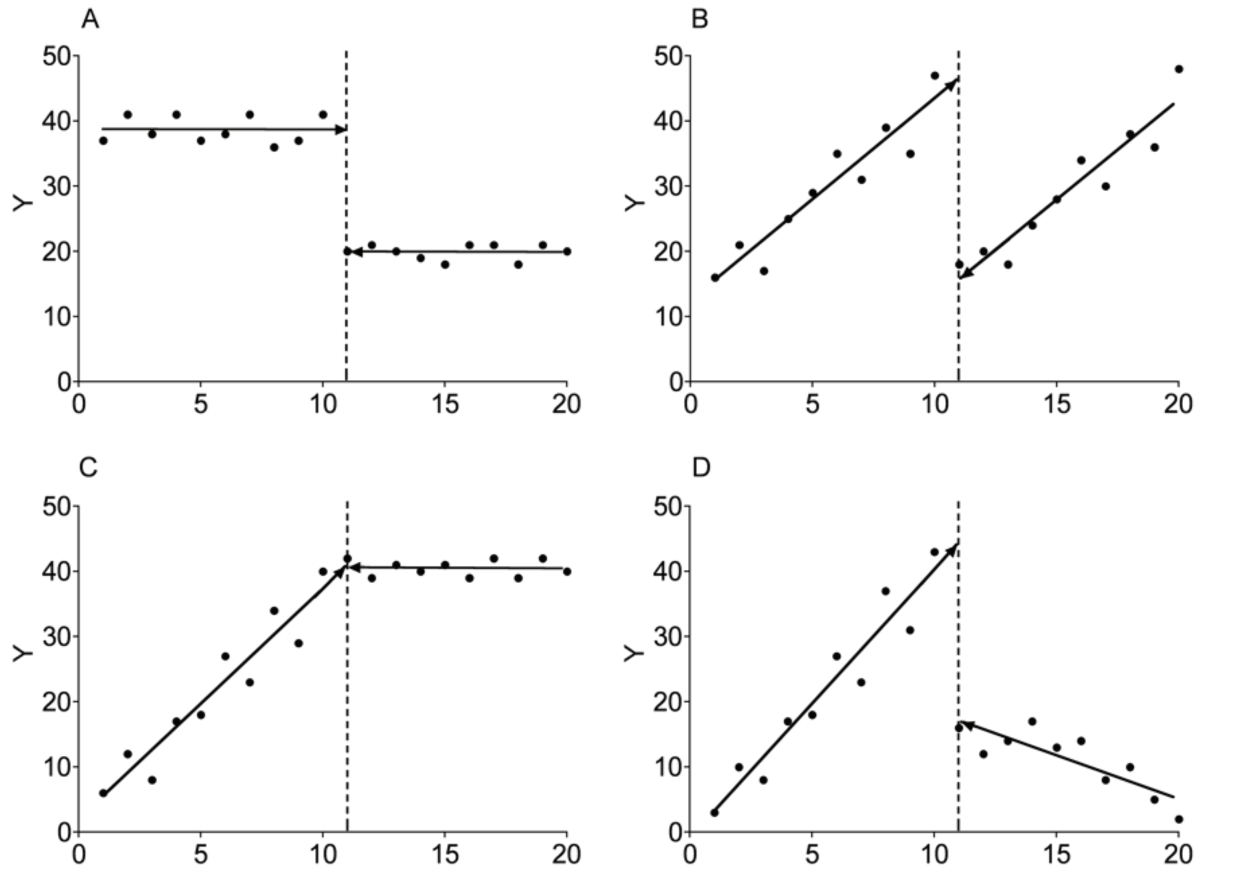


*Figure 1.* Panel A: Perceptual speed data (Holtzman, 1963) that illustrate an apparent change in both level and slope. Panel B: Aphasia data (Robey, et al., 1991) that illustrate an apparent change in both level and slope. Panel C: Weekly diastolic blood pressure readings (Borckardt, 2002) that illustrate little if any change in level and negative change in slope. Panel D: Oral naming accuracy data (Spencer et al., 2000) illustrating a trending series that was not subject to an intervention.

Table 1

Summary of Level Change, Slope Change, and Autocorrelation Estimates Associated with ITSACORR, ITSE, and Three Alternative Methods (H-M OLS, and H-M M-L, and H-M Bootstrap) Applied to Data (Illustrated in Figure 1) from Four Published Sources

| | Method of Analysis | | | | |
|---|---|---|---|---|---|
| | ITSACORR | ITSE | H-M OLS | H-M M-L | H-M Bootstrap |
| Study A | | | | | |
| Level change: | -5.90 | -0.96 | -31.07*** | -30.87*** | -30.61*** |
| Slope change: | -0.87* | -0.99** | -1.01*** | -1.01*** | -1.00*** |
| Autocorrelation: | .68 | .17 | (.15)* | .15 | .22 |
| Study B | | | | | |
| Level change: | 65.91** | 45.13*** | 39.51*** | 40.89*** | 39.71*** |
| Slope change: | 0.74 | 2.98* | 3.65* | 3.73* | 3.48* |
| Autocorrelation: | .54 | .13 | ( -.33) | -.35 | -.18 |
| Study C | | | | | |
| Level change: | -75.28* | -9.14*** | -4.33*** | -2.77 | -2.68 |
| Slope change: | 1.05 | -1.65* | -1.83*** | -1.85*** | -1.96*** |
| Autocorrelation: | -.01 | .56*** | (.51)*** | .61*** | .71* |
| Study D | | | | | |
| Level change: | 55.55*** | 55.55*** | -6.82 | -7.08 | -5.34 |
| Slope change: | -0.24 | -0.25 | -.26 | -.25 | -.26 |
| Autocorrelation: | .12 | -.01 | (-.04) | -.04 | .13 |

*Note*: *$p < .05$; **$p < .01$; ***$p < .001$

## Results

### Inconsistency Between the Logic of the Design and the Estimates Produced by ITSACORR

The intervention effects and autocorrelation estimates associated with ITSACORR, ITSE, and the three methods based on the H-M design matrix appear in Table 1 for the data illustrated in the four panels of Figure 1.

The columns of the table list the methods of analysis and the major rows identify the study; the level change, slope change, and autocorrelation estimates appear in the body of the table.

#### Study A.

The data illustrated in panel A of Figure 1 are perceptual speed measures obtained from a schizophrenic patient each day before and after the administration of chlorpromazine. These data have appeared in publications by several writers (e.g., Crosbie,1995; Glass, et al., 1975; Holtzman, 1963) to illustrate time-series procedures. Crosbie (1995) used these data to support the claim that, in the case of a large number of observations, ITSACORR, ITSE, and ARIMA methods all reach the same conclusion. An examination of Table 1 reveals that ITSACORR and ITSE provided level decrease estimates of 5.90 and 0.96 points, respectively ($p > .50$ for both methods), whereas each of the three remaining methods estimated the level decrease as about 31 points ($p \leq .001$). An ARIMA analysis of these data by Glass, et al. (1975) (not included in Table 1) estimated a drop in level of approximately 22 points ( $p \leq .001$). Visual inspection of the data suggests a level decrease in the neighborhood of 20 - 30 points. All methods included in Table 1 yielded similar slope change estimates. Because the ARIMA model used by Crosbie (1995) as a basis of comparison with ITSACORR and ITSE does not estimate slopes, one could not compare this ARIMA model with the other analyses in terms of slope change. The autocorrelation estimate produced by ITSACORR was a value of .68 while the other procedures yielded autocorrelation estimates that range from .15 to .22.

#### Study B.

The data in panel B appeared in an article by Robey et al. (1999) that strongly promoted the use of ITSACORR. After applying ITSACORR to the data these authors stated that “The t test for a change in level is also significant (i.e., $t = 3.341$, $p = .005$); the t test for a change in slope does not achieve statistical significance (i.e., $t = 0.187$, $p = .855$)” (p. 460). Unfortunately, Robey et al. (1999) did not present the descriptive statistics (i.e., intercept and slope estimates) associated with these t and p values. These descriptive statistics are listed in Table 1.

Notice that ITSACORR estimated the level change as approximately 66 points. If one examines panel B of Figure 1 one can see the elevation of the phase 1 line at time point 9 and the elevation of the phase 2 line for the same time point; it is obvious that they differ by approximately 40 points. Indeed, an inspection of the level change statistic for each analysis shown in Table 1 indicates that only the estimate provided by ITSACORR deviates far from 40 points.

The slope-change and autocorrelation estimates provided by ITSACORR also deviate greatly from the results provided by the other methods. In contrast, all of the other methods provide slope-change estimates that are consistent with the visual impression. Table 1 also shows that ITSACORR provides a higher value for the autocorrelation estimate (i.e., .54) than the estimates provided by the other methods (range = -.35 through .13).

#### Study C.

Borckardt (2002) was written to demonstrate “how clinicians can efficiently conduct scientific analyses of a patient’s response to such interventions using time-series designs supported by newly developed analytic procedures.” (p. 190). One of the analytic procedures to which he referred was ITSACORR. Weekly diastolic blood pressure data from this study appear in panel C of Figure 1. These data were obtained before and after participants received a multimodal psychotherapy intervention. A visual inspection of the data reveals a minor negative slope during the baseline phase, essentially no level change after intervention, and a strong negative shift in slope beginning immediately after the intervention. These visual impressions concur with the results of the statistical methods listed in Table 1, with one exception. ITSACORR estimates a huge decrease in level (over 75 points) and a positive shift in slope. Both of these estimates are grossly inconsistent with the visual appearance of the data. Visual inspection suggests that the drop in level can be no more than a few points. Moreover, as the difference between the minimum and maximum values in the entire series is only 32 points, a level change estimate of 75 points can have no real meaning. The easily discerned visible decrease in slope in the second phase suggests that, even in the absence of supporting statistical evidence (e.g., that produced by the other methods described in Table 1), there is strong reason to question the validity of the positive slope-change estimate produced by ITSACORR. Clearly, the level (or intercept) change, slope change, and autocorrelation estimates associated with the ITSACORR method do not describe these data to any reasonable degree.

Study D.

Spencer et al. (2000) applied ITSACORR to a multiple-baseline design that contained three experimental series and one control series. A visual inspection of their complete data (not illustrated here) reveals a major shift to each experimental phase following the intervention and very little change throughout the control series. Although they did not apply ITSACORR to the control series, such an analysis is illuminating. The control data appear in panel D of Figure 1.

If ITSACORR provides reasonable level change and slope change estimates it should confirm the visual impression of little change in the control series other than an upward trend that is quite consistent throughout the duration of the experiment. Although no intervention interrupted this series, a vertical line was inserted to show the time point at which the intervention interrupted one of the experimental series. As seen in Table 1, the level change estimate provided by ITSACORR is almost 56 points ($p < .001$) even though the intervention was not applied to this series. ITSE yielded essentially the same results. In contrast, the other methods estimate a minor decrease in level that fails to reach statistical significance ($p > .05$). All methods essentially agree with respect to the degree of slope change and autocorrelation.

Summary of Observed Differences Between ITSACORR and Other Methods Regarding Parameter Estimates.

A comparison of the ITSACORR level-change estimates with those provided by three acceptable statistical methods (as well as by visual analysis) reveals major inconsistencies for each published study illustrated in Figure 1. In some cases the ITSACORR estimate approximates the estimate provided by ITSE (an unacceptable method), but often these two methods produce very different estimates. A comparison of results from all analyses reveals level-change estimates for ITSACORR that are as much as 50 times as large as the others. In some cases, the ITSACORR estimate is far larger than the difference between the highest and lowest values in the entire series. Although the discrepancies among level change estimates tended to be larger than the discrepancies among slope change and autocorrelation estimates, discrepancies among the latter measures are also pronounced. Because the results of ITSACORR differ so much from those associated with both visual analysis and acceptable statistical methods it is reasonable to ask why. The next two sections provide answers to this question.

Inconsistency Between the Logic of the Design and the Parameters of the Structural Model

This section focuses on the comparison of the intercept parameters specified in the ITSACORR structural model with the level change parameter dictated by the logic of the two phase design. The ITSACORR structural model [identical to the Gottman (1981) ITSE model] comprises two parts, one for the pre intervention data and one for the post intervention data, as shown below.

ITSACORR Model

Pre intervention (1)

$$Y_t = m_1 t + b_1 + \sum_{i=1}^{P} a_i Y_{t-i} + e_t$$

Post intervention (2)

$$Y_t = m_2 t + b_2 + \sum_{i=1}^{P} a_i Y_{t-i} + e_t$$

where, using Gottman's notation, m1 and m2 are the process slopes for phases 1 and 2, respectively, b1 and b2 are the process intercepts for phases 1 and 2, respectively, ai is the ith autoregressive coefficient, P is the autoregressive order of the model, and et is the error. The time indicator t associated with the outcome variable Y takes on values 1, 2, . . ., n1 for observations in the first phase, and values n1 + 1, . . . , n1 + n2 for observations in the second phase (Gottman, 1981, p. 349). The numbering of the time indicator is crucial in understanding the nature of the intercepts defined for this model.

The difference between the two intercept parameters (i.e. b1 and b2) in this model does not measure the change in level at (or near) the appropriate time point n1 + 1. Both b1 and b2 measure elevation at the time point before the first observation in the first phase (i.e., time period zero). The value of b1 results from extrapolating back only one time point, whereas the value of b2 results from extrapolating from time point n1 + 1 all the way back to time point zero. Although both intercepts are associated with the same time point (i.e., zero), the difference between these two measures does not, in general, yield a measure of level change. One can, however, derive the correct level change parameter from the parameters of the ITSACORR model (Huitema & McKean, 2000a, p. 57). The correct expression for the level change parameter is: (b2 - b1) + (n1 + 1)(m2 - m1). It can be seen from this expression that the intercept difference (b2 - b1) is equivalent to the level change parameter only if the two slopes are exactly the same. Because the intercepts in the ITSACORR structural model define elevation at time period zero rather than time period n1 + 1, the model defines change effects that do not coincide with the logic of the two-phase interrupted time-series design.

Inconsistency Between the Structural Model and the Design Matrix

The first stage in the estimation of the parameters of the ITSACORR structural model can be carried out using the full model ITSE-ITSACORR design matrix shown in the Appendix (panel A) . Nevertheless, this matrix is not consistent with the design matrix that conforms to the structural model. The inconsistency can be seen in the numbering of the time periods for the second phase of the design. The second phase numbering follows the sequence t = n1 + 1, . . . , n1 + n2 in the structural model (presented above), whereas the design matrix actually employed in the ITSACORR analysis (see column four in panel A of the Appendix) uses the sequence t = 1, 2, . . . , n2 . This inconsistency means that the ITSACORR method and the resulting parameter estimates deviate from the ITSACORR structural model (which is also inconsistent with the logic of the design) and the intercept parameters it implies. This distinction between the model and the design matrix serves as an important step in conceptually decomposing the problems with the method.

Unacceptable Inferential Performance

It has been shown that ITSACORR provides unacceptable descriptive results. This outcome eliminates most interest in the inferential aspects of the analysis because there is little reason to consider hypothesis tests (or confidence intervals) applied to invalid parameter estimates. Nevertheless, for the sake of completeness, it is shown in this section that the inferential aspects of the analysis remain invalid even if one ignores the unacceptable descriptive properties of the ITSACORR method.

The inferential approach recommended for ITSACORR comprises a two-stage procedure. First, a preliminary omnibus F-test is carried out to test the following compound hypothesis: H0: m1 = m2 and b1 = b2. This hypothesis states that both slopes are identical and both intercepts are identical for the two phases of the study. The test is based on a comparison of results obtained using the full and reduced model design matrices shown in the Appendix. Rejection of the compound hypothesis is typically interpreted to mean that an intervention effect has occurred in the form of either a slope change or an intercept change (or both). A separate t-test on each sub hypothesis (i.e., H0: m1 = m2 and H0: b1 = b2) is then carried out. Many researchers, however, ignore the preliminary test and attend to only the t's.

At first glance this two stage approach appears consistent with conventional statistical practice outside the time-series context. Upon close inspection, however, it can be seen that the ITSACORR preliminary F-test on the compound hypothesis contains fatal flaws. There has been provided a formal mathematical proof elsewhere (Huitema, McKean, & Laraway, 2007) that illustrates the problem with this test. The essential idea can be conveyed simply. Suppose one has a situation in which there is no level change whatsoever and the slopes are identical (i.e., there is a common slope). As the common slope approaches infinity the difference between ITSACORR intercepts approaches infinity even though the level has not changed. It follows that the difference between intercepts can be infinitely large even though the value of the preliminary F is zero. Because the F-test does not provide information relevant to the evaluation of differences between the intercepts defined for the ITSACORR method, this test has been ignored in the analyses presented in Table 1.

Simulation results regarding the empirical Type I error relevant to the preliminary F-test and the t-tests on change between intercepts and change between slopes are as follows: Type I error for the preliminary omnibus F-test on both intercept and slope change = .25 and .37 when autocorrelation is set at .50 and .80, respectively. The corresponding error rates on the individual test for intercept change equaled .16 and .20, and the corresponding results for the test on slope change equaled .21 and .33. Because the empirical Type I error rates greatly exceed the nominal value the tests do not possess satisfactory inferential properties and the results regarding power are of no interest. Consequently, power results are not provided. Other results, not presented here, show that if realistic levels of slope exist in the first phase, the Type I error rate for the t on intercept change is approximately 1.0.

## Conclusion

The ITSACORR method begins with Gottman's ITSE procedure and adds to it some well-intended modifications. Unfortunately, the descriptive and inferential properties are unacceptable. Each aspect of the whole framework (including the structural model, the design matrix, the autocorrelation estimator, the ultimate parameter estimation scheme, and the inferential method) contains fatal flaws. It can thus be concluded that the ITSACORR method does not provide information that is relevant to the purposes of the interrupted time-series design. Moreover, there is no situation in which one can recommend the use of ITSACORR. This conclusion is clearly at odds with recent recommendations in the literature. Some comments on these published recommendations are in order.

An examination of the foundation supporting the recommendations to use ITSACORR rather than Gottman's ITSE or ARIMA intervention models reveals little more than restatements of claims contained in the original descriptions of the method. Crosbie (1995, p. 391) compared the results produced by ITSACORR with those produced by Gottman's ITSE and an ARIMA moving averages intervention model that Glass et al. (1975) had previously applied to a portion of Holtzman's (1963) perceptual speed data. Crosbie concluded that "all three procedures reach the same conclusion" (p. 392). These methods are not based on the same assumptions regarding the nature of the underlying time-series process and they do not estimate the same parameters. These differences are reflected in the parameters modeled. This is why there are no slopes in the cited ARIMA analysis. Therefore, the claim that ITSACORR, ITSE, and ARIMA procedures "reach the same conclusion" (Crosbie, p. 392) is without foundation. Unfortunately there are several textbooks (e.g., Franklin, Allison, &

Gorman, 1997 and Christensen, 2007) and many recent journal articles that perpetuate this mistaken notion.

Another misunderstanding regarding ITSACORR relative other procedures have recently appeared. Jenson, Clark, Kircher, and Kristjansson (2007) have stated that "ITSACORR yields conservative estimates of intervention effects" (p. 488). Examples presented have been based on published data where this is far from true. Studies C and D in the present article yield ITSACORR estimates of intervention effects that are approximately 10 to 25 times the size of the correct estimates.

Because it has been shown that both the descriptive and inferential properties of ITSACORR are unacceptable it is recommend that this method not be used. More adequate methods include certain ARIMA and regression-based approaches cited in this article; it is recommended that they be given serious consideration when choosing an analysis for interrupted time-series designs.

References

Borckardt, J. J. (2002). Case study examining the efficacy of a multi-modal psychotherapeutic intervention for hypertension. International *Journal of Clinical and Experimental Hypnosis*, *50*, 189-201.

Borckardt, J. J., & Nash, M. R. (2002). How practitioners (and others) can make scientifically viable contributions to clinical-outcome research using the single-case time-series design. *International Journal of Clinical and Experimental Hypnosis*, *50*, 114-148.

Box, G. E. P., & Jenkins, G. M. (1970). *Time-series analysis: Forecasting, and control*. San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. L. (1965). A change in level of a non-stationary time series. *Biometrics*, *52*, 181-192.

Christensen, L. B. (2007). *Experimental Methodology* (10th ed.). Boston: Allyn and Bacon.

Crosbie, J. (1993). Interrupted time-series analysis with brief single-subject data. *J*ournal of Consulting and Clinical Psychology, 61, 966-974.

Crosbie, J. (1995). Interrupted time-series analysis with short series: Why it is problematic; How it can be improved. In J. M. Gottman (Ed.), *The Analysis of Change* (p. 361-395). Mahwah, NJ: Lawrence Erlbaum.

Davidson, K. M., & Tyrer, P. (1996). Cognitive therapy for antisocial and borderline personality disorders: Single case study series. *British Journal of Clinical Psychology*, *35*, 413-429.

Ellis, M.V. (1999). Repeated measures designs. *The Counseling Psychologist*, *27*, 552-578.

Franklin, R. D., Allison, D. B., & Gorman, B. S. (Eds.). (1997). *Design and analysis of single-case research*. Mahwah, NJ: Lawrence Erlbaum.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975*). Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press.

Gorman, B. S., & Allison, D. B. (1997). Statistical alternatives for single-case designs. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and Analysis of Single-case Research* (p. 159-214). Mahwah, NJ: Lawrence Erlbaum.

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. New York: Cambridge University Press.

Gottman, J. M. (Ed.). (1995). *The analysis of change*. Mahwah, NJ: Lawrence Erlbaum.

Gottman, J. M., & Rushe, R. H. (1993). The analysis of change: Issues, fallacies, and new ideas. *Journal of Consulting and Clinical Psychology*, *61*, 907-910.

Hogenraad, R., McKenzie, D. P., & Martindale, C. (1997). The enemy within: Autocorrelation bias in content analysis of narratives. *Computers and the Humanities*, *30*, 433-439.

Holtzman, W. (1963). Statistical models for the study of change in the single case. In C. W. Harris (Ed.) *Problems in Measuring Change* (p. 99-211). Madison: University of Wisconsin Press.

Huitema, B. E. (2004). Analysis of interrupted time-series experiments using ITSE: A critique. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *3*, 27-46.

Huitema, B. E. & McKean, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods*, *3*, 104-116.

Huitema, B. E., & McKean, J. W. (2000a). Design specification issues in time-series intervention models. *Psychological Measurement*, *60*, 38-58.

Huitema, B. E. & McKean, J. W. (2000b). A simple and powerful test for autocorrelated errors in OLS intervention. *Psychological Reports*, *87*, 3-20.

Huitema, B. E. & McKean, J. W. (2007). *Time-Series Intervention Analysis using ITSACORR: Fatal Flaws Expanded version*. Unpublished Manuscript.

Huitema, B. E., McKean, J. W., & McKnight, S. (1994, August). *Small-sample time-series intervention analysis: Problems and solutions*. Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools*, *44*, 483-493.

Lucyshyn, J. M., & Albin, R. W. (1997). Embedding comprehensive behavioral support in family ecology: An experimental, single-case analysis. *Journal of Consulting and Clinical Psychology*, *65*, 241-251.

McCleary, R., & Hay, R. A., Jr. (1980). *Applied time series analysis for the social sciences*. Newbury Park, CA: Sage.

McKnight, S., McKean, J. W., & Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, *5*, 87-101.

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

Rushe, R. H., & Gottman, J. M. (1993). Essentials in the design and analysis of time-series experiments. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Statistical Issues* (p. 493-528). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Spencer, K. A., Doyle, P. J., McNeil, M. R., Wambaugh, J. L., Park, G., & Carroll, B. (2000). Examining the facilitative effects of rhyme in a patient with output lexicon damage. *Aphasiology*, *14*, 567-584.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in education settings. *School Psychology Review*, *26*, 333-368.

Stoline, M. R., Huitema, B. E., & Mitchell, B. (1980). Intervention time-series model with different pre- and post-intervention first-order autoregressive parameters. *Psychological Bulletin*, *88*, 46-53.

Appendix

(A)
ITSE - ITSACORR Full Model Design Matrix (X) and Y Vector

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & Y_1 \\ 1 & 2 & 0 & 0 & Y_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & n_1 - 1 & 0 & 0 & Y_{n_1 - 1} \\ \hline 0 & 0 & 1 & 1 & Y_{n_1} \\ 0 & 0 & 1 & 2 & Y_{n_1 + 1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 1 & n_2 & Y_{N-1} \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} Y_2 \\ Y_3 \\ . \\ . \\ . \\ Y_{n_1} \\ \hline Y_{n_1 + 1} \\ Y_{n_1 + 2} \\ . \\ . \\ . \\ Y_N \end{bmatrix}$$

(B)
ITSE - ITSACORR Reduced Model Design Matrix (XR) and Y Vector

$$\mathbf{X}_R = \begin{bmatrix} 1 & 2 & Y_1 \\ 1 & 3 & Y_2 \\ 1 & 4 & Y_3 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & N & Y_{N-1} \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} Y_2 \\ Y_3 \\ Y_4 \\ . \\ . \\ . \\ Y_N \end{bmatrix}$$