Time Series Analysis in Historiometry: A Comment on Simonton

Wayne F. Velicer
Brett A. Plummer
Cancer Prevention Research Center,
University of Rhode Island

ABSTRACT Time series analysis (TSA) is one of a number of new methods of data analysis appropriate for longitudinal data. Simonton (1998) applied TSA to an analysis of the causal relationship between two types of stress and both the physical and mental health of George III. This innovative application demonstrates both the strengths and weaknesses of time series analysis. Time series is applicable to a unique class of problems, can use information about temporal ordering to make statements about causation, and focuses on patterns of change over time, all strengths of the Simonton study. Time series analysis also suffers from a number of weaknesses, including problems with generalization from a single study, difficulty in obtaining appropriate measures, and problems with accurately identifying the correct model to represent the data. While careful attempts are made to minimize these problems, each is present in the Simonton study, although sometimes in a subtle manner. Changes in how

The number of methods available for data analysis has increased tremendously in the last 25 years. The single item most responsible for the development of these new methods is the computer. Data analysis previously was performed with the aid of such relatively primitive mechanical aids as calculators and slide rules. Almost all methods were univariate and one method in particular, the Analysis of Variance, dominated most published research. The first use of the computer was to provide high-speed computation for existing methods. Then new methods were developed that could only be performed with a computer because they involved complex numerical algorithms. Two characteristics common to many of these new methods were a focus on multivariate procedures and an awareness of the temporal dimension. Some of these procedures, like time series analysis, survival analysis, and structural equation modeling, are becoming well known. A recent book, *Best Methods for Measuring Change* (Collins & Horn, 1991), provides a good overview of many of the new methods for the longitudinal analysis of change. With the increased availability of these new methods, we are reaching the point where we can find a method that will answer our research questions instead of formulating our research questions to fit an available method.

The Simonton (1998) paper provides a clear example of the application of a new data analytic method to a unique problem area. It involves one of the new methods, time series analysis (TSA), applied to a well-studied problem, the recurrent attacks of mental and physical illness experienced by King George III of Great Britain. However, while this problem has long been of interest to historians and psychiatrists, the approach employed in this paper is unique and represents an innovative new approach

Overview: Time Series Analysis

Time series analysis is a statistical procedure appropriate for repeated and equally spaced observations on a single subject or unit. Several texts and recent papers provide good introductions to this method (Box, Jenkins, & Reinsel, 1994; Crosbie, 1993; Glass, Wilson, & Gottman, 1975; Velicer, 1994; Velicer & Colby, 1997; West & Hepworth, 1991). A practical advantage of the procedure is that it is highly appropriate for analyzing types of data available in settings where traditional methods are not appropriate. A theoretical strength is that the method emphasizes the nature of the change process and is appropriate for assessing the pattern of change over time.

Traditional between-groups statistical procedures cannot be employed with this type of data because repeated observations on the same unit cannot be assumed to be independent. The presence of dependency may substantially bias a statistical test that does not take it into account. The direction of the bias will depend on the direction of the dependency. Crosbie (1993) illustrates this bias in a simulation study.

The most widely employed method of analyses for time series designs is based on the ARIMA (Autoregressive Integrated Moving Average) models (Box, Jenkins, & Reinsel, 1994). These procedures permit the effects of dependency to be statistically removed from the data (Glass, Wilson, & Gottman, 1975).

Advantages of Time Series Analysis

Time series analysis has generated widespread interest for a number of

Applicability. It would be difficult to imagine any other type of statistical analysis that would be more appropriate for this data. We have many observations on a single person gathered at equal intervals over a long period of time. Dependency is clearly present in the data, so any method of analysis that does not account for dependency would be inappropriate. This paper represents a clear example of the unique kind of problem that can be studied by TSA.

Causal relationships. The focus of the paper is on the causal relationship between stress and the incidence of madness. An especially unique aspect of time series analysis is the ability to study causal relationships with different temporal relationships, as illustrated by the finding of the 9-month lag. A procedure that did not permit consideration of different time lags between two variables could very well have missed or distorted the type of relationship observed here.

Temporal patterns. The pattern of change over time is central to the analysis. Stress and health both change over time. The pattern of change is not constant. Differences can occur in both the duration and the degree for either stress or health. The central argument of the paper is based on the similarity of the pattern of change over time when lagged by 9 months. The treatment in this case is provided by the natural environment and is not under the control of the researcher.

Drawbacks of Time Series Analysis

The employment of time series methods also suffers from a number of drawbacks. First, generalizability cannot be inferred from a single study.

decline in health after a 9-month lag. Simonton (1998) proposes the reasonable hypothesis that changes in stress level precipitated dietary changes such as increased alcoholic consumption which in turn activated changes in liver function which activated porphyria hepatica. While this represents a reasonable description of a causal mechanism that explains the 9-month lag and is consistent with the known facts, it is also impossible to replicate this study easily, a necessary step to establishing generalizability. This is clearly a major problem with this paper, but it should be more properly recognized as a general problem with this whole research area.

When it is not possible to generalize across subjects, alternative forms of generalizability may be considered. In this case, Simonton (1998) argues from the viewpoint of generalizability across measures, arguing that the results are replicated across various combinations of the pairwise comparisons of the variables. We will discuss a problem with this approach below. For each of the six measures, Simonton (1998) combines all 11 ratings into a single score that can be viewed as an unweighted aggregation. The advantage of combining the 11 ratings is, of course, increased reliability.

An alternative approach would be to treat the individual ratings as replications rather than combining them. We can estimate the reliability of a single rating using the reverse of the usual Spearman-Brown formula. For the total stress rating of .94, the reliability of an individual rating would be, on average, only .59. If this is viewed as unsatisfactory, a compromise might be to create random teams of raters. Assume that we have four teams of 3 raters instead of the current one team of 11 raters. (This is the equivalent of employing item parcels rather than individual

These correlations raise the question of whether we have three separate constructs or three measures of one construct. Since the reported correlations are at the maximum that could be expected based on the reported reliability (following the well-known correction for attenuation logic), it would seem more reasonable to view these as three measures of a single construct and, therefore, perform one analysis on the cross-lagged correlations between the two total scores. This, unfortunately, means that the consistency of the pattern of results across the measures is considerably less impressive. For future studies, greater independence might be achieved by having each measure produced by an independent set of raters. This would avoid the possibility of cross-variable contamination.

Model identification. Model identification remains a troublesome problem in time series analysis. While an extensive variety of procedures has been developed to identify the model (see Velicer & Colby [1997] for a review), no clear consensus as to which method is best has emerged and model identification remains a difficult and problematic procedure. Among the problems are the large number of data points required for accurate identification, the complexity of the procedures, and problems with the accuracy and reliability of some methods, even under ideal circumstances (Velicer & Harrop, 1983). While the number of data points available is adequate here, the method of model identification is not described. Several different models are identified for the different measures but no underlying mechanisms are proposed.

Two different approaches to model identification have developed. In some cases, identification of the model is assumed to be of no intrinsic

1985, 1990). To some extent, this is illustrated by the similarity between the first-differenced and prewhitened analysis (see Table 1, Simonton, 1998).

The other approach to model identification is to identify and attempt to interpret the model, assuming that this will provide information about the basic process underlying the data (see Velicer & Colby, 1997). That was not done in the Simonton (1998) paper, but should have been considered. The most disturbing aspect of the data is that all variables required differencing. This means that none of the six series are stable. While this can be handled statistically by differencing, it would be of interest to try to determine the cause of this problem.

One potential source is the nature of the rating task, since the data were presented in chronological order. The ratings of one month could, therefore, be easily influenced by the rating for the preceding months. This type of contamination could be viewed as temporal contamination. In other words, the instability might be in the rating task and not actually present in the observed events. One solution to this problem would be to have the months presented in a different random order on separate pieces of paper to each rater. This would assure independence from any form of temporal contamination.

Multivariate Time Series Analysis

The analysis performed by Simonton (1998) involved calculating the pairwise cross-lagged correlations between each of the three stress measures and each of the three health measures, for a total of nine sets of cross-lagged correlations. Two other analyses could be considered: a structure of data across time. Dynamic factor analysis extends this method to examine the covariance of any single variable at time t with the covariance of any other variable of higher lags (time t-1, t-2, . . ., t-a). Although complex, the technique is flexible, as factor solutions obtained at different lags may be conceptually different. Factors responsible for short-term behavior may be different than those factors responsible for long-term behavior. Dynamic factor modeling can potentially add a tremendous amount of information to simply observing cross-correlations. Short-term behavior change can be measured not only in terms of measured variables, but in terms of latent constructs as well. One obvious use of dynamic factor modeling is examining the latent structure of variables across time.

Despite the apparent usefulness of the method, dynamic factor analysis is a relatively unknown technique. There are several reasons for this. First, dynamic factor analysis is a new technique (Molenaar, 1985) and its effectiveness and appropriateness across different designs is still being explored. Second, few published papers have employed dynamic factor analysis even as an exploratory technique. Third, few published Monte Carlo studies have examined the magnitude of cross-correlations or the effect that sample size has on making stable parameter estimates. Fourth, little has been written about the statistical assumptions that underlie dynamic factor analysis.

An attempt was made to obtain the dynamic factor solutions for the same data set analyzed by Simonton (1998). We attempted to fit the 3 stress-related variables and the 3 health-related variables with a 2 latent factor model, general stress and general health. Macros provided by Wood and Brown (1994) were used. Although these macros could obtain distinct overall factors of health and stress, although obtaining a two-factor solution would have provided more support for this claim.

Applications of dynamic factor analysis will be most successful when the initial data collection anticipates this type of analysis. If multiple ratings had been used, either from the individual raters or from teams of raters, a latent variable model could have been developed that used the observed variables as the manifest variables. The contribution of the individual raters or teams of raters could have been evaluated. Models that represented alternative theories about how to conceptualize stress could have been compared, including a single latent variable model (total stress), a two latent variable model (personal stress and political stress), and a hierarchical model (two first-order latent variables for the two types of stress and a second-order latent variable representing total stress).

CONCLUSIONS

Simonton (1998) has written an impressive paper that applied an innovative new method of data analysis to a well-studied problem and produced some interesting results. Both the strengths and weaknesses of the paper reflect the strengths and weaknesses of time series analysis. This study represents a novel application, focuses on issues of causality, and describes the pattern of change over time. However, replication is not possible, so the results lack generalizability. The extremely high correlations between the measures suggest that the analysis can be viewed as involving only a single set of cross-lagged correlations rather than nine separate analyses. The presence of instability in all six observed measures is problematic. Changes in how the data are gathered for future

Collins, L. M., & Horn, J. L. (Eds.). (1991). *Best methods for the analysis of change*. Washington, DC: American Psychological Association.

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

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

Harrop, J. W., & Velicer, W. F. (1985). A comparison of three alternative methods of time series model identification. *Multivariate Behavioral Research*, **20**, 27–44.

Harrop, J. W., & Velicer, W. F. (1990). Computer programs for interrupted time series analysis, Part 2: A quantitative evaluation. *Multivariate Behavioral Research*, **25**, 233–249.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL VII user's reference guide*. Chicago: Scientific Software International.

Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time-series. *Psychometrica*, **50**, 181–202.

Molenaar, P.C.M. (1987). Dynamic factor analysis in the frequency domain: Causal modeling of multivariate psychophysiological time series. *Multivariate Behavioral Research*, **22**, 329–353.

Simonton, D. K. (1977). Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin*, **84**, 489–502.

Simonton, D. K. (1998). Mad King George: The impact of personal and political stress on mental and physical health. *Journal of Personality*, this issue.

Velicer, W. F. (1994).Time series models of individual substance abusers. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (NIDA Research Monograph 142), pp. 264–301. Rockville, MD: National Institute on Drug Abuse.

Velicer, W. F., & Colby, S. M. (1997). Time series analysis for prevention and treatment research. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 211–249). Washington, DC: American Psychological Association.

Velicer, W. F., & Harrop, J. W. (1983). The reliability and accuracy of time series model identification. *Evaluation Review*, **7**, 551–569.