Behavior Research Methods, Instruments, & Computers
1987, 19 (4), 407-408

PROGRAM ABSTRACTS/ALGORITHMS

An interactive computer program for assessing serial dependency in longitudinal data and the identification of basic ARIMA models

R. L. BROWN
University of Wisconsin, Madison, Wisconsin

The concern for serial dependency in time-based measures has been voiced by many in the social sciences as a potential threat to the adequate assessment of longitudinal phenomena. This concern has been a point of debate for some time, with one school of thought disregarding the importance of dependency and the other emphatically stating that all longitudinal data are serially dependent. Although some studies do not support the notion of dependency in longitudinal data (Huitema, 1982), the bulk of the literature seems to favor the occurrence of some dependency (Glass, Willson, & Gottman, 1975; Gottman & Glass, 1978; Hartmann et al., 1980; McCleary & Hay, 1980). If dependency exists in longitudinal data, then, as Scheffé (1959) and others (Gastwirth & Rubin, 1971; Padia, 1973) have demonstrated, estimates of Type I errors using tests based on independent error distributions will be underestimated.

Dependency in longitudinal data may be viewed as the t^{th} data value (z_t) predicted to some degree from previous values ($z_{t-1}, z_{t-2} \ldots z_{t-p}$) or from previous residuals or random shocks ($a_{t-1}, a_{t-2} \ldots a_{t-q}$.) If dependency occurs in this fashion, we find a correspondence to two acceptable statistical processes and, subsequently, a possible solution to the problem of dependency in the analysis of longitudinal data. The two statistical processes are called the autoregressive (AR) and the moving average (MA) processes (Box & Jenkins, 1970), sometimes symbolized as ARIMA (p,d,q) for AutoRegressive Integrated Moving Average process, where p = the number of AR parameters, d = the number of differences needed for stationarity, and q = the number of MA parameters.

The autoregressive process (ARIMA $(p,0,0)$) involves the prediction of the value z_t from previous values of z_{t-i}, subsequently regressing the series of values upon itself i values in the past. This process assumes that no value of z is made initially at $t=0$, and thus the level of the series value (L) is an unknown. This process may be expressed as z deviations from L:

$$z_t L = \phi_1(z_{t-1}-L)+\phi_2(z_{t-2}-L)+\phi_p(z_{t-p}-L)+a_t,$$

where ϕ_p = the autoregressive parameter, and a_t = random shock (residual), with a_t = NID(0, σ^2).

The moving average process (ARIMA $(0,0,q)$), on the other hand, involves the concept of a moving average of random shocks (a_t), where z_t may be regarded as being dependent on the current random shock (a_t) and a portion of the previous random shock (a_{t-q}). The moving average process may be expressed as:

$$z_t - L = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_q a_{t-q},$$

where θ_q = the moving average parameter.

The use of ARIMA modeling is governed by an iterative process of (1) model identification, (2) parameter estimation, and (3) model diagnosis. Although many very good routines are available for parameter estimation, one finds little assistance in the area of dependency recognition and model identification, the initial step in the iterative modeling process. To date, only one major program has been designed to help users identify ARIMA models, the AUTOBOX program, and its subset program AUTOBJ (Reilly, 1986). Although AUTOBOX seems to be an excellent program, its cost places it out of reach of many (Shumway, 1986).

This paper describes a reduced form of the AUTOBOX/AUTOBJ program, which provides users with assistance in the identification of appropriate ARIMA models and the diagnosis of the model's residuals. It must be noted that parameter estimation must be conducted outside of the TIMEID program. For example, one may easily use Minitab-PC (Ryan, Joiner, & Ryan, 1986) for passing files and estimating parameters.

Program

The TIMEID program has been developed from an earlier interactive mainframe teaching program that helped users identify ARIMA models (Brown, 1983). TIMEID is written and compiled in QuickBasic and, as currently configured, operates only on IBM PC machines with a minimum of 256K, using DOS 3.2. This restriction is due to a machine-contingent utility embedded within TIMEID. Since it is not feasible for me to provide a version for each machine configuration, I decided to configure the current version for IBM equipment with the latest DOS version. The program can be made to work on all IBM PC compatible machines by obtaining a public-domain screen utility program called Menu Maker (Fluegelman, 1984). This program can be obtained through local computer bulletin board systems, or it can be requested with the TIMEID program disk. Instructions for machine conversion are provided on the TIMEID program disk.

Input and Output. Since the program was designed both for research and classroom use, it has been made

Portions of this work were supported in part by National Institute of Mental Health Grant 1 R01 MH40886-01 to Mary Ann Test. Address correspondence to R. L. Brown, Schizophrenia Research Unit, 352A/425 Henry Mall, University of Wisconsin, Madison, WI 53706.

Table 1
TIMEID Model Identification Performance

Sample Size	Proposed Model	Identified Model	Index of Identification*
40	ARIMA(0,0,1)	ARIMA(0,0,1)	19%
40	ARIMA(1,0,0)	ARIMA(1,0,0)	27%
46	ARIMA(0,1,1)	ARIMA(1,0,0)	16%
—	—	ARIMA(0,1,1)†	18%
100	ARIMA(0,1,1)	ARIMA(0,1,1)	18%
71	ARIMA(2,0,0)	ARIMA(2,0,0)	26%
60	ARIMA(0,1,1)	ARIMA(0,1,1)	29%

*This is a measure of fit comparing the calculated AUCFs and PACFs to theoretical values in accordance with basic ARIMA models. The larger this percentage, the less confidence one should place in the identified model. †Once the series was differenced, TIMEID correctly identified the model.

user friendly by incorporating multiple window overlays. The program calculates zero, first, second, and third differenced autocorrelations and partial autocorrelations on data sets as large as 200 observations. In addition to providing graphic representations of the autocorrelation (AUCF) and partial autocorrelation (PACF) functions, the program also provides examples of theoretical correlograms for a variety of nonseasonal nonmixed models. Routines for graphing and subsetting raw data sets are also available. Probably the most interesting facet of the program is its ability to automatically provide the user with an assessment of data dependency and an identification of a probable ARIMA model. The algorithm that provides the automatic identification (upon request) is a simple expanding exponential function based on the calculated AUCF and PACF for the data series (for details see Brown, 1982). Although the algorithm is simplistic, tests indicate that it does an adequate job in identifying basic ARIMA models. Table 1 shows six previously identified models and the results of TIMEID's automatic identification.

The TIMED program did quite well and provided only one incorrect identification. This particular model was considered a borderline model; once differenced, TIMEID identified the model correctly.

Program Availability. The program and a brief user's guide are available from R. L. Brown, Schizophrenia Research Unit, 352A/425 Henry Mall, University of Wisconsin, Madison, WI 53706. Send a self-addressed mailer with sufficient postage and a 5.25-in. floppy disk for a copy of the TIMEID program disk.

REFERENCES

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

Brown, R. L. (1982, August). *The use of computer-assisted identification of ARIMA time-series*. Paper presented at the meeting of the American Psychological Association, Washington, DC.

Brown, R. L. (1983). TIMCAI: An interactive program to assist in the identification of univariate nonseasonal nonmixed ARIMA time series. *Behavior Research Methods & Instrumentation*, **15**, 101-102.

Fluegelman, A. (1984). *Menu maker* [Computer program]. Minneapolis, MN:Headlands Press.

Gastwirth, J. L., & Rubin, H. (1971). Effect of dependence on the level of some one-sample tests. *Journal of the American Statistical Association*, **66**, 816-820.

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

Gottman, J. M., & Glass, G. V. (1978). Time-series analysis of interrupted time-series experiments. In T. Kratochwill (Ed.), *Single subject research*. New York: Academic Press.

Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. S. (1980). Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis*, **13**, 543-559.

Huitema, B. E. (1982, August). *Autocorrelation in behavioral data: A methodological artifact*. Paper presented at the meeting of the American Psychological Association, Washington, DC.

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

Padia, W. L. (1973). *Effect of autocorrelation on probability statements about the mean*. Unpublished master's thesis, University of Colorado, Boulder.

Reilly, D. P. (1986). *AUTOBOX: The user's guide*. Hatboro, PA: Automatic Forecasting Systems.

Ryan, T. A., Jr., Joiner, B. L., & Ryan, B. F. (1986). *Minitab reference manual*. University Park, PA: Pennsylvania State University, Minitab Project.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Shumway, R. H. (1986). AUTOBOX version 1.02. *The American Statistician*, **40**, 299-301.

(Revision accepted for publication May 29, 1987.)