DOCUMENT RESUME

ED 110 463 TM 004 703

AUTHOR Padia, William L.
TITLE Model Identification in Time-Series Analysis: Some Empirical Results.
PUB DATE [Apr 75]
NOTE 4p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Evaluation Methods; Identification; Item Analysis; *Mathematical Models; Observation; Research Design; *Research Methodology; Research Problems; Statistics; Test Construction; *Test Results; *Time

ABSTRACT

Model identification of time-series data is essential to valid statistical tests of intervention effects. Model identification is, at best, inexact in the social and behavioral sciences where one is often confronted with small numbers of observations. These problems are discussed, and the results of independent identifications of 130 social and behavioral time-series by two judges are presented. The majority (75 percent) of the series were represented by one of four basic models: "white noise" (i.e., independent observations); first-order autoregressive; first-order moving averages model in the first difference; and "white noise" in the first difference. (Author)

AERA '75 Time-Series

Model Identification in Time-Series Analysis: Some Empirical Results

William L. Padia
Laboratory of Educational Research
School of Education
University of Colorado
Boulder, Colorado 80302

The interrupted time-series experimental design has recently become a useful methodology in the social sciences. Campbell and Stanley (1963) sensitized social scientists to the many possibilities for the design but much of the theoretical framework remained to be discovered prior to formal application. Box and Tiao (1965) and Box and Jenkins (1970) developed the basic statistical structure and Glass, Willson and Gottman (1972) developed the statistical machinery to test for intervention effects in time-series studies.

The "success" of the application of these statistical methods depends on the proper identification of the underlying model assumed for the data. Unfortunately, model identification remains a rather inexact pursuit. The implications of incorrectly identifying a model are, in many cases, a serious threat to the validity of the results. The purpose of this paper is to present the results of the independent identification of some 130 time-series by two judges and to focus on some relevant aspects of the model identification problem.

Most of the data were taken from research reports in social or behavioral publications. The series are not random samples; they represent personal choices of one judge accumulated over the last several years. However, the series do reflect a variety of things observed (e.g., a person, a city, a nation) and a range of applications: alpha brain waves, crime rates, examination scores, stock prices, word association test scores, students' time spent studying, learning curves, etc. The length of the series varied from less than 20 to over 200 time points.

Each judge identified the series by observing the correlogram or pattern of autocorrelations and partial autocorrelations. The degree of differencing d necessary to produce stationarity was first chosen and then the appropriate p and q were selected to complete the identification of the series as an Auto-Regressive-Integrated-Moving-Average model of order p, d and q (ARIMA(p,d,q)). Agreement was obtained on approximately 80% of the series with much of the 20% disagreement due to inherent model ambiguity. Consensus was reached with respect to the disagreements and the series were categorized according to the value of p, d and q.

Approximately 75% of the non-seasonal series (18% of the series were identified as seasonal) were covered by four models: The "white noise" process ARIMA(0,0,0), 18%; the first-order autoregressive process ARIMA(1,0,0), 23%; "white noise" after first differencing ARIMA(0,1,0), 11%; the integrated moving-averages process ARIMA(0,1,1), 23%. In no case was differencing above the second order required to produce stationarity. The most often encountered models were the first-order autoregressive (23%) and the first-order integrated moving averages.

The process of identifying these series brought to light rather clearly some of the problem areas regarding proper choice of p, d and q. The salient components of model identification are the following.

Identifying d. This is a critical area since over-differencing a series introduces dependency and under-differencing fails to remove dependency. In the series identified about half were $d=0$ and half were $d=1$ with only 6 cases $d=2$.

Equivalence of Models. While unique model representation is desirable, there are inherent ambiguities among the various possibilities. For example, the ARIMA(1,0,1) with autoregressive parameter phi of one is formally equivalent to the ARIMA(0,1,1). In general, the choice of representation depends on the values of the autoregressive and moving-average parameters phi and theta.

Deterministic Drift. Series with a consistent rise or fall can be identified by inspecting the mean of the dth difference. A statistically significant non-zero mean implies the need for incorporation of a drift paramenter into the model formulation. Alternatively, the raw series may be adjusted to some base rate to correct for the drift (e.g., traffic fatality data adjusted by constant miles driven).

Seasonal Model. Seasonal models are identifiable by large autocorrelations at lag s, the length of the cycle. The series may be identified as seasonal and represented as a complex multiplicative model (e.g., (0,0,1)x(0,1,1)) or by adjusting the original series for cyclic fluctuations and then identifying the adjusted series.

Changes in the Model. Series in the social and behavioral science areas seem to be uniquely beset with problems of the intervention altering the fundamental nature of the series. For example, the Scandanavian countries' traffic series show marked discrepancies between the pre and post-WWII periods. Proper identification of both pre and post-intervention series is essential for correct formulation into the statistical estimation phase of the intervention effect.

While these problems appear formidable they can, for the most part, be reduced in complexity since most of the series actually encountered in practice are rather simple in form.

REFERENCES

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

Box,G.E.P. & Tiao,G.C. A change in level of a non-stationary time-series. Biometrika, 1965,52, 181-192.

Campbell,D.T. & Stanley,J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1966.(Also appears as Ch. 5, "Experimental and quasi-experimental designs for research on teaching," in Gage,N.L.(editor), Handbook of research on teaching. Chicago: Rand McNally, 1963.)

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