DOCUMENT RESUME

ED 135 859 TM 006 090

AUTHOR Padia, William L.
TITLE The Consequences of Model Misidentification in the Interrupted Time-Series Experiment.
PUB DATE [Apr 77]
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Error Patterns; *Hypothesis Testing; *Mathematical Models; Probability; Research Design; Research Methodology; Social Science Research; *Statistical Analysis; Tests of Significance; Validity
IDENTIFIERS *Interrupted Time Series; Time Series Analysis; Type I Error

ABSTRACT

Campbell (1969) argued for the interrupted time-series experiment as a useful methodology for testing intervention effects in the social sciences. The validity of the statistical hypothesis testing of time-series, is, however, dependent upon the proper identification of the underlying stochastic nature of the data. Several types of model misidentifications are examined for some commonly encountered models in the social sciences. Analytic expressions for actual Type I error and power probabilities are derived when the mathematics is tractable; simulation techniques are adopted for the remainder of the cases. Results indicate that model misidentification leads to severe perturbations of the nominal probabilities. (Author)

The Consequences of Model Misidentification in the Interrupted Time-Series Experiment

WILLIAM L. PADIA

California Department of Education

Campbell (1969) argued for the interrupted time-series experiment as a useful methodological scheme for testing intervention effects in the social sciences. The validity of the statistical hypothesis testing of time-series is, however, dependent upon the proper identification of the underlying stochastic nature of the data. Several types of model misidentifications are examined for some commonly encountered models in the social sciences. Analytic expressions for actual Type I error and power probabilities are derived when the mathematics is tractable; simulation techniques are adopted for the remainder of the cases. Results indicate that model misidentification leads to severe perturbations of the nominal probabilities.

THE CONSEQUENCES OF MODEL MISIDENTIFICATION IN THE INTERRUPTED TIME-SERIES EXPERIMENT [1]

William L. Padia
Office of Program Evaluation and Research
California Department of Education
721 Capitol Mall
Sacramento, California 95814

The interrupted time-series experiment is a useful methodological scheme for testing intervention effects in the social sciences. The wide range of application, spurred by Campbell (1969), both in the experimental and quasi-experimental settings, has produced a number of interesting designs for testing complex interventions.

The design and analysis settings are well explicated by Glass, Willson and Gottman (1975) and a user-oriented computer program by Bower, Padia and Glass (1974) is available to simplify the analysis. Certainly one of the most troublesome and amorphous aspects of time-series methodology is in area of model identification. With the high degree of ambiguity present in the model identification stage, it is necessary to know the results of mis-classification errors. The social scientist must be aware of the severity and direction of these errors.

The stochastic properties of time-series following the Box and Jenkins (1970) formulation, and the importance of these properties as related to the significance testing of intervention effects are considered below. The salient problem area in regard to statistical hypothesis testing is the proper identification of the underlying stochastic process of the time-series. Improper model identification yields spurious Type I error and power estimates, the nature of which requires systematic examination.

---

[1] A paper presented at the annual meeting of the American Educational Research Association in New York, April, 1977.

The most common types of model misidentifications are based on common time-series models observed in social and behavioral settings, Padia (1975): white noise processes in undifferenced data; white noise processes in first-differenced data; first-order autoregressive processes; and, integrated moving average processes. Misidentification is defined to encompass the following three areas:

1. Underfitting one model to another such as identifying a second-order moving average process as a first-order moving average process.
2. Misfitting one model to another such as identifying a first-order autoregressive process as a first-order moving average process.
3. Improper specification of d. The difference parameter d may be underestimated as in identifying an integrated moving average process as a moving average process in the zero$^{\underline{th}}$ difference. Similarly, d may be overestimated as in identifying a first-order autoregressive process as a first-order autoregressive process in the first difference.

Combinations of these three areas are also considered. Analytic expressions for actual error probabilities under the various misidentifications are derived when the mathematics is tractable; Monte-Carlo simulation techniques are adopted for the remainder of the cases. The results are valid for series in which the number of observations is large: analytic solutions are accurate to order $1/n$ and simulation results are based on series of length n=122. Detailed analytic solutions and simulation procedures found in Padia (1975).

The salient point with regard to the various misidentifications is that the mis-specification of d is the dominant influence and, for the most part, controls the direction and magnitude of discrepancies from the nominal Type I error or the nominal power. Failure to identify d properly leads to spurious error probabilities and the direction and magnitude is predictable, in general, irrespective of the stochastic properties or parameter values of the models. Proper specification of d is imperative in any statistical significance testing procedure involving time-series data.

Several conclusions may be drawn based on a summary of the results with some extrapolations where the regularity of the results justifies generalization. All conclusions refer to model misidentifications for certain parameter values that are likely to be observed under a particular model misidentification.

Underfits

Underfits produce Type I error rates which are functions of the magnitude and sign of the autoregressive and moving-average parameters. The effects are mixed in that some combinations of parameters produce actual Type I probabilities less than the nominal, while other combinations produce α rates greater than the nominal. The results are valid regardless of the value of the difference parameter d, as long as there is agreement with d for the misidentified model and the true model. For the most part, the deviations from the nominal α level are not trivial and care should be taken to avoid an underfit. Overfitting is never a problem since the additional overfit parameters are zero in the true model.

Misfits

Misfits, like underfits, produce true Type I error rates which are functions of the magnitude and sign of the true and misidentified models. The effects are mixed and the results are identical for various values of d, as long as d in the assumed model agrees with d in the true model.

$\underline{d}$ Underestimations

The most spectacular effects on Type I error occur when a process which is non-stationary in level or non-stationary in level and slope is assumed to be stationary at $\underline{d} = 0$. The perturbed Type I error probabilities are 0.80 to 0.90, quite independent of the stochastic properties of the misidentified models. The probablistic "wanderings" of a series non-stationary in level nearly guarantees the existence of a significant intervention effect under a true null hypothesis. Indeed, the effects are so severe that the actual rates are very nearly identical for all three nominal α levels of 0.10, 0.05 and 0.01.

With a series non-stationary in level and slope which has been identified as non-stationary in level only, the situation is quite different. In the case of a white noise underfit of a white noise process, the nominal α overestimates the actual α, contrary to the $\underline{d} = 1$ underfits. Once the effects of a series non-stationary in level are accounted for (by first differencing), the various models and parameter values are free to exert their influence on the nominal Type I error probabilities. The deviations from nominal remain serious with the actual α less than nominal α in some cases, and greater in other cases.

The effects of underestimating $\underline{d}$ within a given model are not additive, i.e., actual Type I probabilities for a $\underline{d} = 0$ to $\underline{d} = 2$ underfit are not the sum of a $\underline{d} = 0$ to $\underline{d} = 1$ underfit and a $\underline{d} = 1$ to $\underline{d} = 2$ underfit.

$\underline{d}$ Overestimations

Overestimating $\underline{d}$ is not as serious as underestimating $\underline{d}$. Overdifferencing a $\underline{d} = 0$ series once produces mild discrepancies in the white noise cases with actual α less than the nominal α. The effect of a $\underline{d} = 1$ to $\underline{d} = 0$ overestimate for models other than white noise, generally results in actual α less than nominal α, with the different model parameters exerting limited influence.

Overdifferencing a $\underline{d} = 1$ process yields actual α levels greater than nominal α levels in the white noise case and, for the most part, in models other than white noise. The type of stochastic process which is overdifferenced and the values of the parameters have some effect as to the direction and magnitude of the discrepancies.

The effects of overdifferencing are not additive, i.e., actual Type I error probabilities obtained by overdifferencing a $\underline{d} = 0$ process twice are not the sum of the probabilities associated with overdifferencing a $\underline{d} = 0$ and $\underline{d} = 1$ process.

Combinations

When a combination of a misfit or underfit with a mis-specification of $\underline{d}$ occurs, the resulting effect on Type I error is a function of the $\underline{d}$ mis-specification. With $\underline{d}$ underestimates (true $\underline{d} = 1$) the effects of a misfit or underfit are completely obfuscated by the overwhelming influence of the $\underline{d}$ mis-specification. The probabilities are similar to the $\underline{d}$ underestimates considered alone, i.e., in the 0.80 to 0.90 range. With $\underline{d}$ underestimates (assumed $\underline{d} = 1$, true $\underline{d} = 2$) the departures from nominal are not as spectacular, the effect of the underestimate is not as dominate (although, in general, actual α is less than nominal α as the underestimate predicts) and the model and model parameters have some influence.

As expected, in the case of a $\underline{d}$ overestimate (assumed $\underline{d} = 1$, true $\underline{d} = 0$ or assumed $\underline{d} = 2$, true $\underline{d} = 1$) in combination with a misfit or underfit, the dominant influence is the mis-specification of $\underline{d}$. The actual α is close to the nominal α, although the models and their parameters have some influence as to the direction and magnitude of the discrepancy.

Since the predominant influence on the system is the mis-specification of d, the effects on Type I error on the combined misidentification model are not the sum of the separate effects, but controlled rather tightly by the mis-specification of d.

Power Conclusions

Since power and Type I error form an inextricable relationship, the effects of model misidentification on power are closely related to the effects of model misidentification on Type I error. In those cases where d is properly identified, the nominal power may overestimate or underestimate the actual power; the extent and direction of the deviation from nominal is a function of both the particular misfit or underfit and the values of the model parameters. The departures from nominal are moderate to severe depending upon the particular case under study.

In those situations where d is underestimated (true d = 1), the result is that an intervention effect will be detected regardless of its magnitude, the true power being near to 0.9 for the three α levels considered. The probabilities of detecting *any* intervention effect are nearly identical to the probabilities of detecting an intervention effect in the absence of one (Type I error). In other words, the stochastic fluctuations of a non-stationary series insure that any non-zero intervention is detectable with high probability. When a series is identified as non-stationary in level when it is non-stationary in level and slope, the results are that nominal power is an overestimate of the actual power for white noise cases, and that the actual power is nearly identical to the Type I error rate under a true null. As in the Type I error cases, the various models and their parameter values are more important in their influence on power.

In situations where $\underline{d}$ is overestimated, the result is to grossly overestimate the power. Overdifferencing a series removes the intervention effect from all but a few points (depending on the degree of over-differencing), and reduces power to the level of Type I error under a true null.

Once again, the combination misidentification cases reflect the importance of the $\underline{d}$ mis-specification. When a $\underline{d} = 1$ process is underestimated (assumed $\underline{d} = 0$), the power to detect a non-zero intervention effect is at least 0.80, the nominal power greatly underestimating the actual power. When a $\underline{d} = 2$ process is underestimated (assumed $\underline{d} = 1$) and in those cases where $\underline{d}$ is overestimated, the nominal power overestimates the actual power, consistent with the effects of the mis-specifications considered alone. In almost all cases the power is approximately the same as the probability of a Type I error under a true null hypothesis. Power results for the combined models are <u>not</u> the sum of the component power results.

In conclusion, it is obvious that the proper identification of $\underline{d}$ is the highest priority of model identification: an error at this stage produces severe departures from nominal probabilities of significance testing. Proper identification of $\underline{d}$ is a necessary but not sufficient condition for the researcher to operate at nominal probability levels. Once $\underline{d}$ is properly identified, other misidentifications are possible, creating mildly discrepant to severe departures from the nominal probability values depending upon the particular underfit or misfit.

Some general conclusions are warranted based upon the frequency of observed models in the social sciences. Since second differencing is rarely encountered (less than seven percent of the time), these statements apply to series where the difference parameter is zero or one.

In general, underdifferencing places the researcher in the position of greatly underestimating power and operating very liberally with respect to Type I error; overdifferencing produces situations where the researcher overestimates power and operates conservatively with respect to Type I error. Misidentifications where d is properly identified lead to either conservative or liberal situations, depending on the values of the model parameters.

The researcher has two courses of action to minimize the chances of a model misidentification. The first is to select or generate time-series with a large number of observations (50 pre-I and post-I as a minimum). With a smaller number of points it is difficult to determine the degree of differencing required to attain stationarity, since a "wandering" over the short run may be either highly autocorrelated stochastic fluctuations of a stationary series or the "drifting" of a non-stationary series.

In addition to identifying large series, the researcher should always entertain more than one model formulation and examine the residuals for the presence of autocorrelation. A statistical test is available (Box and Jenkins, 1970, p. 503) and has been incorporated into a computer program developed by Bower, Padia and Glass (1974).

Certainly, the researcher can minimize the possibility of model misidentification and the severe accompanying effects on error probability statements by following the above precautions and through a knowledge of the data.

REFERENCES

Box, G.E.P., and Jenkins, G.M. _Time-Series Analysis: forecasting and control_. San Francisco: Holden Day, 1970.

Bower, C.P.; Padia, W.L.; Glass, G.V. TMS: two fortran IV programs for analysis of time-series experiments. Boulder: Laboratory of Educational Research, 1974.

Campbell, D.T. Reforms as experiments. _American Psychologist_, 1969, _24_, 409-429.

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

Padia, W.L. Model identification in time-series analysis: some empirical results. Presented at the 1975 American Educational Research Association, Washington, D.C., 1975.

Padia, W.L. The consequences of model misidentification in the interrupted time-series experiment. Ph.D. dissertation, University of Colorado, July 1975.