DOCUMENT RESUME

ED 090 248 TM 003 512

AUTHOR Willson, Victor L.
TITLE ANCOVA Procedures in Time-Series Experiments: An Illustrative Example.
PUB DATE Apr 74
NOTE 9p.; Paper presented at the American Educational Research Association Annual Meeting (Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Analysis of Covariance; Intervention; *Models; *Research Design; *Statistical Analysis; Traffic Accidents
IDENTIFIERS *Time Series Experiments

ABSTRACT

A statistical model for analysis of multiple time-series observation is briefly outlined. The model incorporates a change parameter corresponding to intervention or interruption of the dependent series. The additional time-series are included in the model as covariates. The practical application of the procedure is illustrated with traffic fatality data from several Northeast states. The intervention variable is a legislative crackdown on speeding offenders in Connecticut, which was claimed to have produced a significant drop in the traffic fatality rate. The analysis using neighboring states as covariates results in a change to nonsignificance for the intervention effect. The seasonal nature of the data also illustrates the applicability of the procedure when seasonality is present, so that no season adjustments are required. (Author)

ANCOVA Procedures in Time-Series Experiments: An Illustrative Example[1]

Victor L. Willson

University of Minnesota

The quasi-experimental time-series design noted by Campbell and Stanley (1963) has been examined by a series of authors in the last few years. Box and Tiao (1965) derived a statistical method for estimating intervention effects in time-series using a discrete stochastic model, the integrated-moving averages (IMA) process.[2] Campbell and Ross (1968) discussed such an analysis applied to time-series of traffic fatalities in Connecticut between 1951 and 1960. Glass, Willson, and Gottman (1972) considered the methodology to date and produced an initial attempt to organize the extant work on the time-series experiment problem. The present paper reports some additional theoretical work and provides a methodological example in which some time-series variables act as covariates on a dependent variable.

Box and Tiao (1965) utilized the simplest non-stating stochastic IMA process, IMA (1,1):

$$z_t - z_{t-1} = -\theta a_{t-1} + a_t, \tag{1}$$

where $\underline{z_t}$ is an observation of the dependent variable at time t, $\underline{a_t}$ is an unobserved random variable uncorrelated over time, and $\underline{\theta}$ is a coefficient of autocorrelation which takes values between -1 and 1. Nonzero values of $\underline{\theta}$

---

[1]Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April 19, 1974.

[2]See Box and Jenkins (1970).

produce the characteristic autocorrelation of $\underline{z}$ found in time-series. If observations of $\underline{z}$ were not correlated, ordinary parametric statistics could be used to evaluate interventions in a time-series. Scheffe (1959) and others have pointed out the serious effects of autocorrelation upon sampling distributions of t-statistics, making usual normal parametric tests unusable. Box and Tiao (1965) derived a method for transforming away the autocorrelation. They noted that the model in (1), may be rewritten as

$$z_t = L + (1 - \theta) \sum_{i=1}^{t-1} a_i + a_t, \tag{2}$$

where L is a location parameter for the first observation (time-series of this sort do not have expected values). The variable $\underline{z}$ was shown to be transformable to $\underline{y}$,

$$y_t = z_t - z_{t-1} + \theta y_{t-1}, \tag{3}$$

which is in the form of the general linear model (let $-\theta = 1 - \lambda$),

$$y_t = (1 - \lambda)^{t-1} L + a_t, \tag{4}$$

or

$$y_t = X\beta + a. \tag{5}$$

With suitable distributional assumptions about $\underline{a}$ (normal, independent and identically distributed with finite variance), the $\underline{\beta}$ is estimable in terms of a design matrix X and observation $\underline{Y}$ under least-squares estimation. If the intervention of the time-series experiment, δ, is included with equation (2) after n_1 observations,

$$z_t = L + \lambda \sum_{i=1}^{t-1} a_i + \delta + a_t, \tag{6}$$

Then equation (4) becomes

$$y_t = (1 - \lambda)^{t-1} L + (1 - \lambda)^{t-n_1-1} \delta + a_t, \tag{7}$$

and both L and δ may be estimated by method of least-squares. The value for λ has not been specified. Box and Tiao (1965) suggested that an iterative method be used (calculate estimates for value of λ between the theoretical limits 0 and 2), using a minimum variance criteria or a Bayesian maximum likelihood function on the residual variance.

Campbell and Ross (1968) discussed the time-series analysis of data on traffic fatalities in Connecticut between 1951 and 1960. An intervention effect was hypothesized to have occurred on January 1, 1956, based upon a "crackdown" on speeding offenders ordered by Governor Abraham Ribicoff of the State of Connecticut on December 23, 1955. The Governor's action required suspension of driving licenses for persons convicted of speeding offenses. First offenders lost their licenses for thirty days, second offenders lost theirs for sixty days, and third offenders had their licenses suspended indefinitely. Inspection of the traffic fatality rate per 100,000,000 driver miles shows a large drop in the rate between December, 1955, and January, 1956 figures. This was cited by the Connecticut State Administration as evidence of the effectiveness of the "crackdown" action.

Analysis of the Connecticut data in which seasonal variations were removed showed a significant difference from zero for the hypothesized intervention. Similar data for nearby states--Massachusetts, Rhode Island, New York, and New Jersey--were examined, and similar drops in level were noted in Massachusetts, New York, and New Jersey. This evidence was presented as disconfirmation of the hypothesis that the crackdown produced a significant drop in the level of Connecticut fatalities per 100,000,000 driver miles (see Glass, 1968).

Glass, Willson, and Gottman (1972) noted that the analysis of multiple time-series may be considered a multiple-group, multiple-intervention design. A multiple comparisons analysis of the estimated intervention effects observed in the five states was presented. A planned orthogonal comparison between the Connecticut change in level and the mean of the changes in level for the other four states results in a t-statistic whose probability of occurrence by chance is .074. The data were seasonally adjusted by subtracting the monthly average for ten years from each month's value.

Several difficulties appear in this analysis. First, it is not clear that the same problem is being addressed when seasonal adjustments are made. Although traditional in economics, the series observed after adjustment are not the same as those observed before in terms of λ (see Willson, 1972). Also, the common design question arises: does one assign a variable as a factor in the design or covary on the variable? Initial differences on the variable are cited as a reason for ANCOVA. Willson (1973) derived a statistical treatment for covariation in time-series. The basic assumption is that the random variable $\underline{a}$ is not truly random but is correlated with the covariate $\underline{c}$. This may be checked statistically. If significant correlation is observed, $\underline{a}$ is decomposed into two components,

$$a_t = \phi(c_t - \bar{c}) + e_t, \tag{8}$$

when $\bar{c}$ is the observed mean of $\underline{c}$ prior to intervention, and in which $\underline{e}$ is now identically, independently distributed (normality is assumed for estimation of confidence intervals). The result is that instead of equation (7),

$$y_t = (1 - \lambda)^{t-1} L + (1 - \lambda)^{t-n_1-1} \delta + \phi(c_t - \bar{c}) + e_t, \tag{9}$$

so that now L, δ, and ϕ are estimated. Multiple covariates may be added much as in multiple ANCOVA. With this procedure, the Campbell and Ross data were reanalyzed as an ANCOVA problem.

The Rhode Island data are ignored since a good case may be made for noncomparability of population and miles driven with respect to the other states, resulting in unstable estimates for the states. The original data are examined without removal of the seasonal variation. The data for three states are graphed in Figure 1.[3] The seasonal components are hypothesized not to affect estimations since regression coefficients and the least-squares procedures are based on deviations from the mean level, which will be matched for different variables on the seasonal component.

The intercorrelation matrix of lag zero first differences[4] is presented in Table 2 for the raw data on fatality rates per 100,000,000 driver miles for the four states under consideration. Zero lag is chosen since it may be assumed that the causes of traffic fatalities operate simultaneously for all the states in a similar fashion.

TABLE 1: Intercorrelation Matrix of Fatalities Per 100,000,000 Driver Miles for Four States for Lag Zero First Difference Data.

	Connecticut	Massachusetts	New York	New Jersey
Connecticut	1.00	.22	.18	.38
Massachusetts	.22	1.00	.15	.22
New York	.18	.15	1.00	.56
New Jersey	.38	.22	.56	1.00

[3]New York was ignored; explanation is given in the next paragraph.

[4]Variables may be regressed on time-lagged data. Here the correlations showed no significant lagged correlations so that regression was contemporaneous in time. First differences ($z_t - z_{t-1}$) are calculated because the model is an IMA (1,1) (Campbell and Ross, 1968).

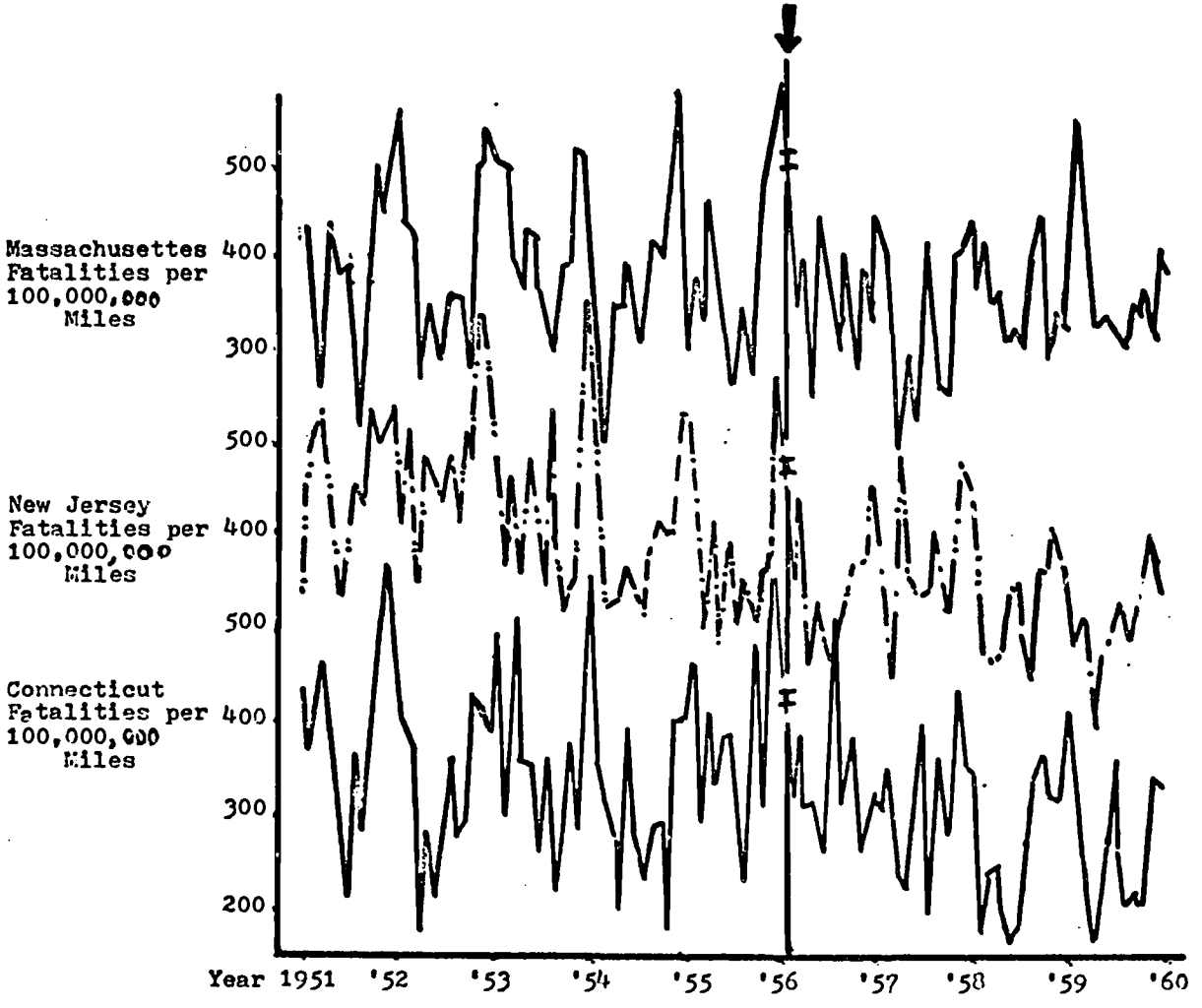

Figure 1 : Graphs of Traffic Fatalities per 100,000 Driver Miles in Massachusettes, New Jersey, and Connecticut for 1951 to 1960 (data from Glass, 1968).

TABLE 2 : Estimates of Parameters for Analysis of Connecticut Traffic Fatalities per 100,000,000 Driver Miles With and Without Massachusetts and New Jersey Traffic Fatalities per 100,000,000 Driver Miles as Covariates.

Pre-Intervention Data: 60 Observations by Month (January, 1951-December, 1955).
Post-Intervention Data: 48 Observations by Month (January, 1956-December, 1960).

With Covariate	Without Covariate
$\hat{\lambda}_z = 0$	$\hat{\lambda}_z = 0$
$\hat{L} = 3.39$	$\hat{L} = 3.57$
$SE(\hat{L}) = .11$	$SE(\hat{L}) = .12$
$\hat{\delta} = -.22$	$\hat{\delta} = -.61^*$
$SE(\hat{\delta}) = .17$	$SE(\hat{\delta}) = .18$
$\hat{\sigma}^2_{Conn} = .622$	$\hat{\sigma}^2_{Conn} = .831$
$\hat{\sigma}^2_{Mass} = .790$	
$\hat{\sigma}^2_{N.J.} = .864$	
$\hat{\phi}_{Mass} = .399^*$	
$SE(\hat{\phi}_{Mass}) = .099$	
$\hat{\phi}_{N.J.} = .254^*$	
$SE(\hat{\phi}_{N.J.}) = .106$	

*($p < .05$ for significance from zero).

The variables associated with New Jersey and Massachusetts are most highly correlated with the Connecticut variable, and the two are chosen for the analysis as covariates.[5] The estimates of parameters with and without covariates are given in Table 2.

Comparison of the mean levels with and without covariates shows that the two estimates are similar in value. The change in level due to the Ribicoff intervention is significantly different from zero in the Connecticut data without covariates, but the intervention is reduced to a nonsignificance when covariates are included. The regression coefficients for New Jersey and Massachusetts fatality rates are both significantly different from zero ($\rho < .05$) when the covariates are included in the analysis.

The main point is that the intervention is reduced to nonsignificance with the inclusion of the covariates. This is the same conclusion reached by Campbell and Ross (1968), and by Glass (1968), but the conclusion is based upon simultaneous mathematical analysis of the four-variable system (independent, dependent, and two covariates) rather than upon separate analyses. This procedure seems preferable to an approach like multiple comparisons analysis when the units of observation cannot be selected randomly. This parallels the desirability of using analysis of covariance over multifacet analysis of variance when randomization is not possible and groups may be initially different. Also, no manipulation of the series in terms of seasonal adjustment is necessary.

---

[5]As in ANCOVA, one desires covariates highly correlated with the dependent variable and not with each other.

References

Box, G. E. P. and Jenkins, G. M. Time-Series Analysis: Forecasting and Control. San Francisco: Holden Day, 1970.

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

Campbell, D. T. and Ross, H. L. The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 1968, 3, 33-53.

Glass, G. V. Analysis of data on the Connecticut speeding crackdown as a time-series quasi-experiment. Law and Society Review, 1968, 3, 55-76.

Glass, G. V., Willson, V. L, and Gottman, J. M. Design and Analysis of Time-Series Experiments. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1972.

Scheffé, H. The Analysis of Variance. New York: John Wiley & Sons, 1959.

Willson, V. L. Concomitant Variation in the Interrupted Time-Series Experiment. Doctoral dissertation, University of Colorado, August, 1973.

Willson, V. L. Estimation of intervention effects in seasonal time-series. Research Paper No. 51. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1972.