AN ANALYSIS OF COVARIANCE MODEL FOR INTERRUPTED TIME SERIES EXPERIMENTS

Bradley E. Huitema, Western Michigan University

I. Introduction

The time series quasi experiment is often a useful design in cases where randomization is impossible but data can be collected across time. If there are n_1 observation points before an intervention to the time series process and n_2 observation points after the intervention, there is generally interest in analyzing pre-post changes in the process. Among the methods of analysis available for this interrupted time series design are those suggested by Box and Tiao (1965 and 1975), Glass, Willson and Gottman (1975) and Jones, Crowell and Kapuniai (1969).

A basic problem with this design is that events concomitant with the planned intervention must be considered as alternative explanations of the change. One method of dealing with this interpretation problem is to employ one or more control series in which the intervention is not applied. If the change in the control series is not the same as the change in the basic series, evidence for the effect of the intervention is strengthened.

A method of analyzing change in the basic series which is free, in a linear sense, of the change in one or more concomitant series is described in this paper. The procedure involves

(1) regressing the basic time series on the concomitant series for the preintervention data,

(2) fitting a first order autoregressive (Markov) model to the residuals of (1) and

(3) testing differences in the postintervention phase between observed points and points predicted from information contained in (a) the concomitant series and (b) autoregression in the residuals of the fitted regression.

Certain aspects of this procedure are extensions of the Jones model.

II. The Model

The proposed model for the time series process is $Y_T = \mu_Y + \beta_{1\cdot 2,3,\cdots,m}(X_{1,T} - \mu_{X_1}) + \beta_{2\cdot 1,3,\cdots,m}(X_{2,T} - \mu_{X_2})\cdots$
$+ \beta_{m\cdot 1,2,\cdots,m-1}(X_{m,T} - \mu_{X_m}) +$
$\alpha\left[Y_{T-1} - (Y|X_1,X_2,\cdots,X_m)_{T-1}\right] + \varepsilon_T$

where

Y_T is the dependent variable score at time T which is any of the equally spaced observation points,

μ_Y is the process mean for the basic (dependent variable) series,

μ_{X_1} through μ_{X_m}, are the means of the concomitant series (i.e., covariates) one through m,

$\beta_{1\cdot 2,3\cdots m}$ through $\beta_{m\cdot 1,2,\cdots,m-1}$ are the partial regression coefficients obtained from regressing Y_T on covariates $X_{1,T}, X_{2,T},\cdots,X_{m,T}$

$X_{1,T}, X_{2,T},\cdots,X_{m,T}$ are scores on covariates $1,2,\cdots,m$ measured at time T. These scores are obtained from any available set of concomitant time series and may be in the form of continuous scores or dummy values which indicate the presence or absence of a condition.

α is the first order autoregression parameter relating the residuals $\left[Y_{T-1} - (Y|X_1,Y_2,\cdots,X_m)_{T-1}\right]$ and $\left[Y_T - (Y|X_1,X_2,\cdots X_m)_T\right]$ and

ε_T is the error which is NID(o,σ^2).

III. Estimation

The parameters of the model are estimated as follows:

μ_Y is estimated using

$$\hat{\mu} = \frac{1}{n}\sum_{T=1}^{n_1} Y_T$$

where n_1 is the number of preintervention observations,

μ_{X_1} through μ_{X_m} are estimated using

$$\hat{\mu}_{X_i} = \frac{1}{n_1}\sum_{T=1}^{n_1} X_i,$$

the partial regression coefficients are estimated using ordinary least squares on the n_1 preintervention points and the first order autoregression parameter is estimated using

$$\hat{\alpha} = \frac{SP_1}{SS_c} =$$

$$\frac{\sum_{T=2}^{n_1}\left[Y_{T-1} - (Y|X_1,X_2,\cdots,X_m)_{T-1}\right]\left[Y_T - (Y|X_1,X_2,\cdots X_m)_T\right]}{\left(\frac{n_1-1}{n_1}\right)\sum_{T=1}^{n_1}\left[Y_T - (Y|X_1,X_2,\cdots X_m\right]_T^2}$$

where SP_1 is the lag one sum of products of the residuals of the regression and SS_c is the zero lag sum of squares corrected by a term that allows for the difference between the number of observations that are associated with the sum of products and sum of squares.

IV. Testing for Intervention Effects

Two related tests are suggested for testing for change in the time series following the intervention.

A. Testing for Postintervention Change at Individual Postintervention Points

The test statistic for evaluating the change in the time series at a post intervention point specified *a priori* is

$$t = \frac{Y_T - \hat{Y}_T}{\sqrt{\left[\frac{SS_c\ (1-R^2-\hat{\alpha}(1-R^2))}{n_1-m-3}\right]\left[1 + \underline{c}_T'(\underline{X}'\underline{X})^{-1}\underline{c}_T\right]}}$$

where

$\hat{Y}$ is the predicted postintervention value based on the fitted model,

R^2 is the coefficient of multiple determination based on the fitted model,

$\underline{c}_T$ is the unity augmented column vector of covariate scores measured at time T, i.e.,

$$\underline{c}_T = \begin{bmatrix} 1 \\ X_{1,T} \\ X_{2,T} \\ \cdot \\ \cdot \\ \cdot \\ X_{m,T} \end{bmatrix},$$

$\underline{X}$ is the unity augmented covariate score matrix based on the n_1 preintervention data points, i.e.,

$$\underline{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \cdot & \cdot & \cdot & X_{m,1} \\ 1 & X_{1,2} & X_{2,2} & & & & X_{m,2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1,n_1} & X_{2,n_1} & \cdot & \cdot & \cdot & X_{m,n_1} \end{bmatrix}$$

The test statistic t is compared with the critical value of the conventional t statistic based on n_1-m-3 degrees of freedom.

B. Testing for Overall Change in the Whole Postintervention Series OVERALL CHANGE ACROSS ALL

If interest lies in evaluating the intervention points, the following approximate test is suggested

$$z = \frac{\sum_{n_1+1}^{n_1+n_2} t_T}{\sqrt{n_2 \sqrt{(n_1-m-3)/(n_1-m-5)}}} .$$

If the first order autoregressive model fits the residuals of the regression, the individual t tests will be approximately independent and the test statistic z will be approximately a standard normal variable.

V. Example

Drunkenness arrest data from two Michigan counties are plotted in Figures A and B. A program that was expected to have an effect on the arrests in the first county (Kalamazoo) is the intervention that occurs after week 39. A comparison of the data from the experimental county with the covariate data from the control county (Calhoun) which was not exposed to the program, reveals a somewhat disturbing pattern. The arrests appear to drop for both the experimental and control counties. In order to evaluate whether or not the postintervention change is significant for the experimental county after controlling for change in the control county, we apply the tests of Section IV.

Individual tests:

Week	t
40	-1.48
41	-1.41
42	- .63
43	-1.45
44	-1.11
45	-1.74
46	- .63
47	-1.25

Overall test:

$$z = \frac{-9.70}{\sqrt{8(35/33)}} = -3.33.$$

None of the individual t tests are significant using $\alpha = .05$, but the observed values associated with these tests are all less than the predicted values. This is indicated by the negative signs associated with the t values. As would be expected, when the combined information from these individual t values is employed in the overall test, the conclusion is that significant postintervention change, beyond that which is found in the control county, took place in the experimental county.

References

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

Box, G. E. P., and Tiao, G. C. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 1975, 70, 70-79.

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

Jones, R. H., Crowell, D. H., and Kapuniai, L. E. Change detection model for serially correlated data. Psychological Bulletin, 1969, 71, 352-358.

Figure A: Kalamazoo County Weekly Total Arrests For Drunkenness

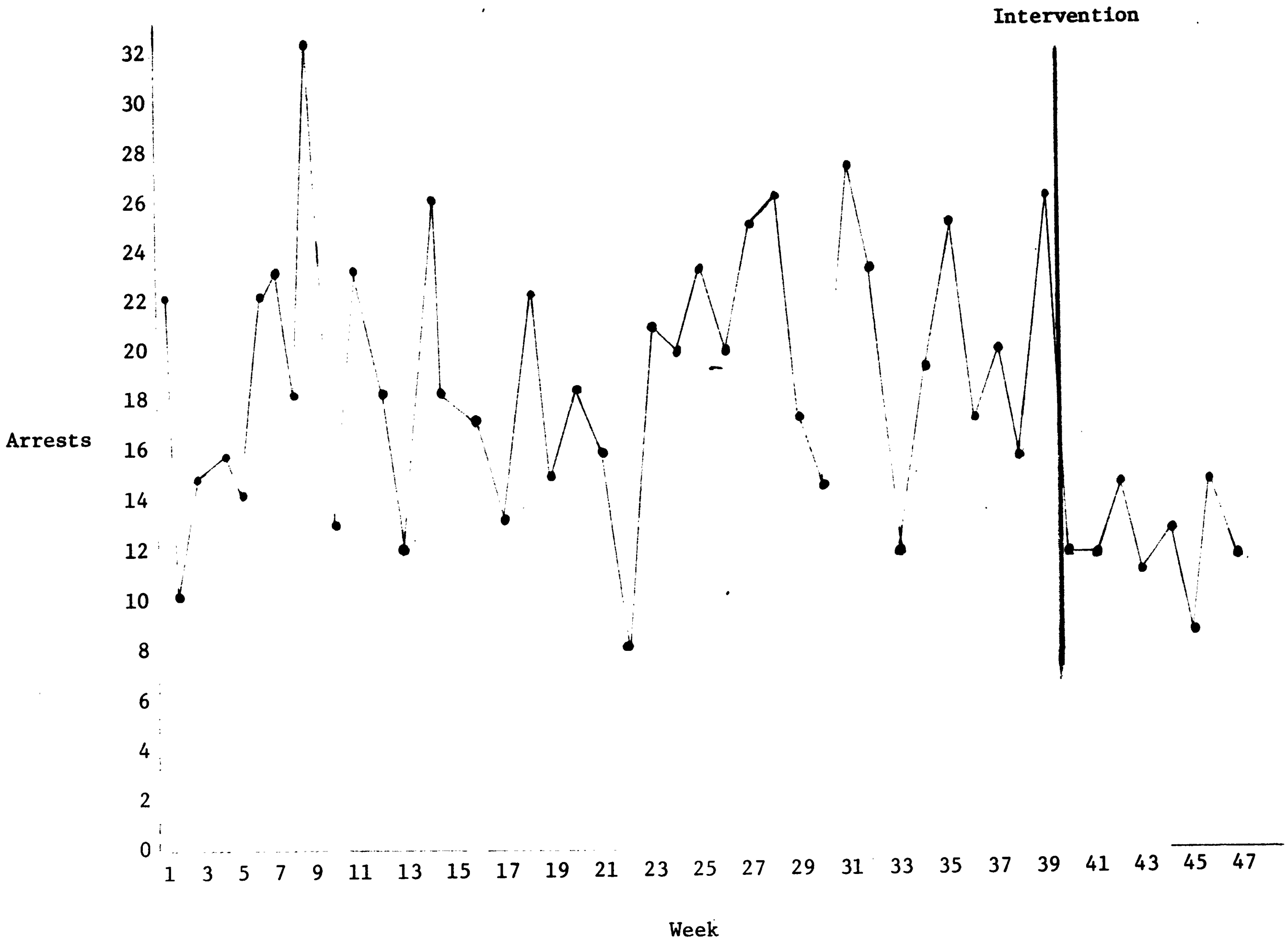

Figure B: Calhoun County Weekly Total Arrests For Drunkenness