seem unreasonable to call for more frequent incorporation of the time dimension in future social science research.

NOTE

1. Another alternative is to use a cross-sectional nonrecursive causal (or simultaneous equation) model, in which X is viewed as both a cause and an effect of Y. However, to obtain meaningful coefficient estimates for such a model, several strict assumptions—including error terms that are completely uncorrelated with each other—must be met. Carmines discusses nonrecursive models very briefly in Chapter 2, and Berry (1984) offers a more exhaustive treatment.

references

Berry, William D. (1984) Nonrecursive Causal Models. Beverly Hills, CA: Sage.

Box, George E. P. and Gwilym M. Jenkins (1976) Time Series Analysis. San Francisco: Holden-Day.

Eulan, Heinz and Michael Lewis-Beck (1985) Economic Conditions and Electoral Outcomes: United States and Western Europe. New York: Agathen Press.

9

Interrupted Time Series

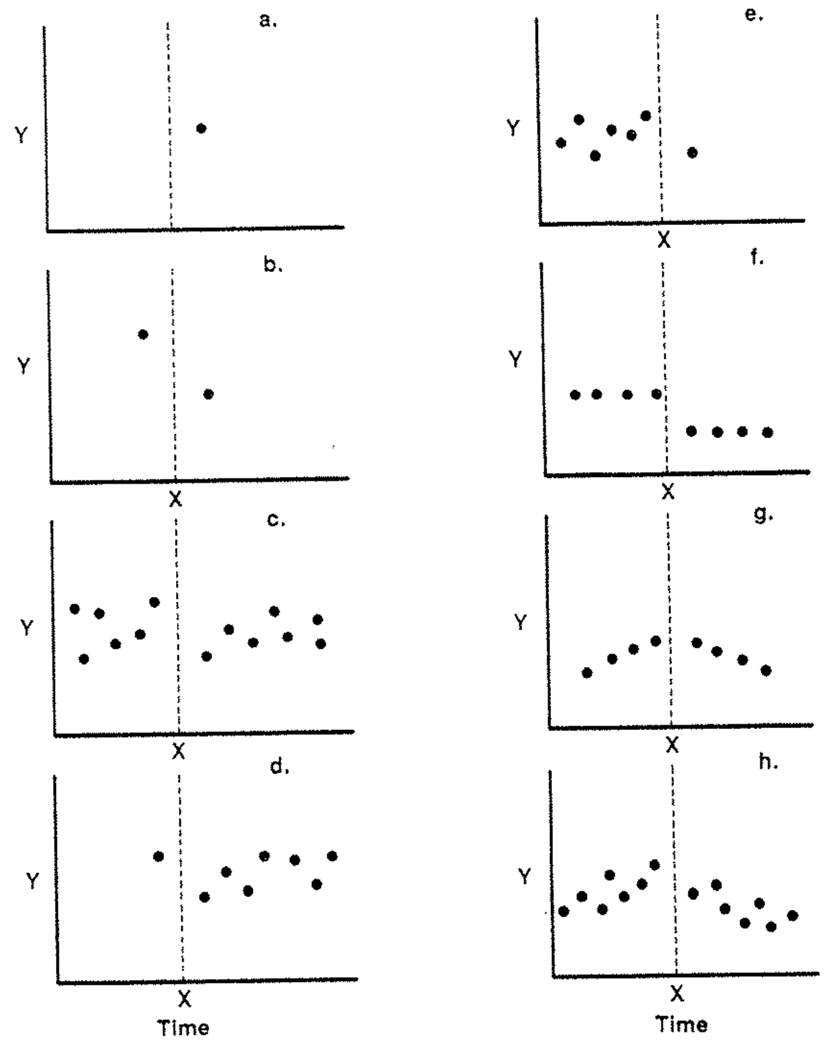
MICHAEL S. LEWIS-BECK

Interrupted time-series (ITS) analyses are becoming rather widely applied in social science research. The technique attempts to answer a simple question: Does the occurrence of a particular event change a variable's behavior over time? The substantive literature offers several examples that pursue this inquiry. An ITS approach has been employed as follows: (1) to assess the impact of new public policies, in areas such as social security, coal mine safety, and civil rights (Albritton, 1979; Lewis-Beck and Alford, 1980; McCrone and Hardy, 1978); (2) to evaluate the effects of specific laws, such as traffic and gun control measures (Campbell and Ross, 1968; Glass, 1968; Deutsch and Alt, 1977; Zimring, 1975); and, (3) to determine the influence of major political shifts, including reform and revolution (Campbell, 1969; Caporaso and Pelowski, 1971; Lewis-Beck, 1979). With respect to the methodology of ITS analysis, work has tended to emphasize the broad issue of research design or the narrow issue of modeling the autocorrelation process (see, in order, Campbell and Stanley, 1966; Campbell, 1969; Cook and Campbell, 1979; Hibbs, 1977; Box and Tiao, 1975; McDowall et al., 1980). Although this effort benefits from both of these past emphases, the focus here is different from either. Simply put, I view ITS models as an interesting variation on the classic multiple regression model, which can be estimated with ordinary least squares, once the standard assumptions are satisfied. Below, I first consider how much can be learned about the effect of an "interruption" merely by "eyeballing" the time series. Then, I present a straightforward ordinary least-squares method for assessing more precisely the effects of both simple and multiple interruptions. Next, I discuss how to bolster confidence in the causal inferences made from these regression estimates, examining in particular the assumptions of no specification error, no measurement error, no perfect multicollinearity and, finally, no autocorrelation. Throughout, in order to make the presentation concrete, I draw on data examples.

EYEBALLING

An ITS analysis should be used whenever the investigator is interested in the effect of a relatively discrete event on a phenomenon observed over time. Normally, the research question is whether or not the event (X) caused a change in the phenomenon (Y). Occasionally, a glance at the scatterplot provides at least a preliminary answer. However, some plots are easier to decipher than others. Different possibilities will be evaluated, as they are sketched in Figure 9.1, where "X" marks the time the event took place. Further, to make the example less abstract, suppose the research question concerns the impact on traffic fatalities (Y) of a hypothetical 1972 reduction (X) in the Kentucky speed limit for automobiles to 50 mph (the example is inspired, of course, by Glass, 1968; Campbell and Ross, 1968). As can be seen immediately, there exists a class of plots (9.1a to 9.1e) that reveal little, if anything, about a cause-and-effect relationship. The first, Figure 9.1a, merely looks at fatalities the year after the new speed limit was imposed, completely ignoring what happened before. Although it is perhaps of interest as a "case study," this single post-event observation really suggests nothing about the effect of the speeding reduction. The second, Figure 9.1b, which records a pre-event speeding observation as well, points to some influence from the 1972 speeding crackdown. However, this influence cannot be asserted with any confidence, for either observation could be a poor representative of Kentucky driving patterns. In other words, because we are not able to rule out the possibility that one or both of the observations are "outliers," causal inference is exceedingly weak.

In Figure 9.1c, the additional data allow the conclusion that observations made regarding 1971 and 1973 are probably not outliers. Still, it is far from clear that the speeding crackdown had any impact. Is the post-event scatter of points generally lower than the pre-event scatter, or is it the same height? Visually, one cannot be sure, because the data are so "noisy." For all these Figures (9.1a–9.1c), inference is difficult because the observations are so few. With ITS analysis, then, as with time-series analysis generally, the more cases the better. But, with ITS analysis an additional concern emerges, which is that enough observations be available before and after the event. For instance, in Figure 9.1d, there are several observations after 1972, but only one before. Given the almost total lack of prior data, it is quite difficult to assess whether the fatality rate actually changed after the speed limit imposition. The same holds true for Figure 9.1e; with no more than one observation after the event,



NOTE: X = the 1972 speed limit change; Y = annual traffic fatalities.

Figure 9.1 a-h: Various Interrupted Time Series Plots

change assessment is highly problematic, despite the relatively large number of pre-event observations.

From Figures 9.1a–9.1e, one sees that the number of observations in general, and their placement to the left and right sides of the event in particular, are important for easy visual diagnosis of the impact of an intervention. But, even a scarcity of cases does not automatically prevent accurate eyeballing, provided the pattern of effect is sharp. For example, in Figure 9.1f, there are only eight observations, but they form two dis-tinct plateaus before and after the event, which clearly indicates a down-ward shift in the level of fatalities subsequent to the speeding crackdown. Again, in Figure 9.1g, there are few observations, but those before form a rising line, and those after form a declining line, suggesting a change in the trend in fatalities as a result of the 1972 law. Of course, the plots in Figures 9.1f and 9.1g are idealized. Time series of observations from the real world never unerringly track straight lines, so we are always eager for more cases in order to better discern what linearity might be present. The scatterplot in Figure 9.1h offers a more realistic picture of such data. Here, the level of fatalities appears to have dropped immedi-ately after the speeding crackdown. Further, the trend in the fatality rate has changed, and is now heading downward. From just eyeballing the scatterplot of Figure 9.1h, it seems that the 1972 speed limit altered the Kentucky traffic fatality rate for the better, both in the short and long run. Exactly how much did it change? Did it really even change at all? Because these questions require answers more precise than eyeballing allows, I turn now to the estimation of ITS models.

ESTIMATION

The basic interrupted time-series model postulates intervention-induced changes in the level (mean) and/or trend (slope) of a time series. The least-squares procedure for estimating an ITS equation is straight-forward, involving the use of dummy variables to represent the interven-tion (see also Draper and Smith, 1966: 134–142). First, I look at the simple ITS model (SITS), which aims to estimate the impact of a single interven-tion. As an example, I explore the effect of the Cuban revolution on Cu-ban economic growth. Then, I go on to the multiple ITS model (MITS), which estimates the impact of multiple interventions. That example ex-amines the impact of the various U.S. coal mine safety laws on the mine fatality rate. After presenting these estimation procedures, I turn to the problems of making causal inferences from these ordinary least-squares (OLS) results.

The most uncomplicated SITS model evaluates the change in the level of the time series after the intervention occurs. One such change in level is diagrammed in Figure 9.2a, where the series jumps to a new plateau

following the intervention. The data evidence no trend, and Y_t is a perfect function of the intervention, yielding the following regression equation:

$$Y_t = b_0 + b_1 (D_{1t}) \quad [9.1]$$

where Y_t = the value of the dependent variable at time point t; D_{1t} = a dichotomous dummy variable scored 0 for those observations before the event and 1 for those observations after the event, and data are available for N time points denoted 1, 2, . . . N. For the preintervention period, then, equation 9.1 for Y_t simplifies to

$$Y_t = b_0 + b_1 (0) \quad [9.2]$$

$$Y_t = b_0 \quad [9.3]$$

— Before Intervention

but for the postintervention period,

$$Y_t = b_0 + b_1(1) \quad [9.4]$$

$$Y_t = b_0 + b_1 \quad [9.5]$$

— After Intervention

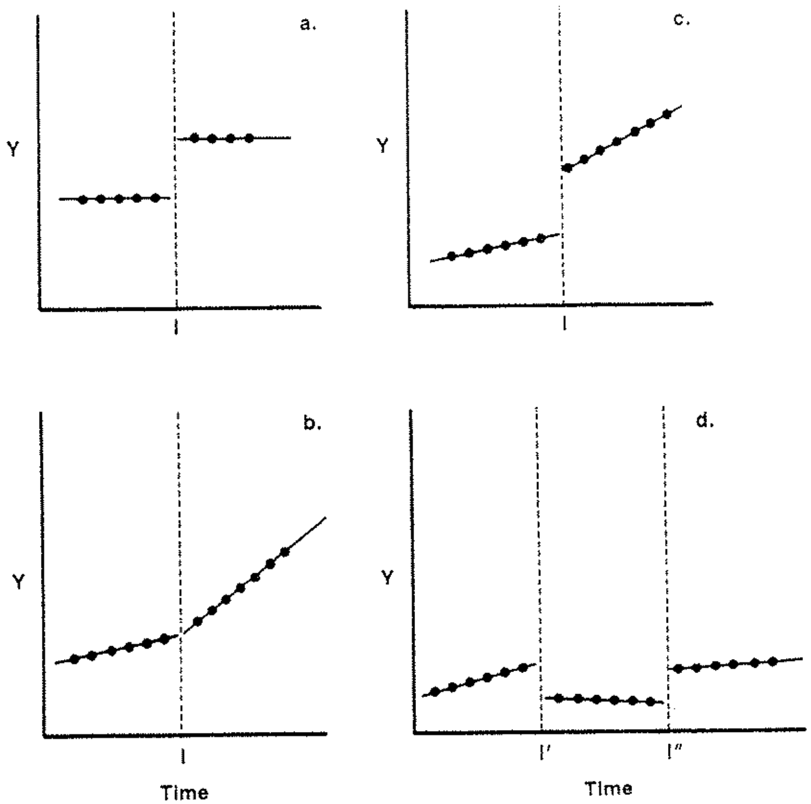
Hence the bivariate regression equation 9.1 estimates the change in the level (the change in the mean value) of the series, which is "$+b_1$". Observe that this change in level is actually a change in the intercept of the prediction equation, from b_0 to $b_0 + b_1$.

The Figure 9.2a diagram is unrealistic in several respects, one of which is its lack of any trend. Time-series data often exhibit a trend, which may obscure the change in level, and which may even change it-self. In Figure 9.2b, the trend in the time series changes after an interven-tion. This postintervention change in trend can be estimated with the following multiple regression equation:

$$Y_t = b_0 + b_1 X_{1t} + b_2 D_{2t} \quad [9.6]$$

where Y_t = N time-series observations on the dependent variable; X_{1t} = a dummy variable counter for time from 1 to N; D_{2t} = a dummy variable counter of time scored 0 for observations before the event and 1, 2, 3 . . . for observations after the event. For the preintervention pe-riod, then, the prediction equation simplifies to

$$Y_t = b_0 + b_1 X_{1t} + (0) \quad [9.7]$$



NOTE: Y = the dependent variable; I = the point of intervention.

Figure 9.2 a-d: Some Idealized Interrupted Time Series Plots

$$Y_t = b_0 + b_1 X_{1t} \quad \text{Pre} \qquad [9.8]$$

Thus the slope (b_1) estimates the trend before the intervention. For the postintervention period, however, the prediction equation does not so simplify, because the values of D_{2t} increase incrementally with time. The trend for the postintervention period, then, is actually $b_1 + b_2$. Therefore, b_2 estimates the change in the trend (slope) of the time series subsequent to the intervention.

The more general SITS model allows for the possibility of estimating simultaneously both an intercept change and a slope change. (An example in which both occur is depicted in Figure 9.2c.) The appropriate multiple regression equation for estimating these changes is the following:

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} \quad \text{Post} \qquad [9.9]$$

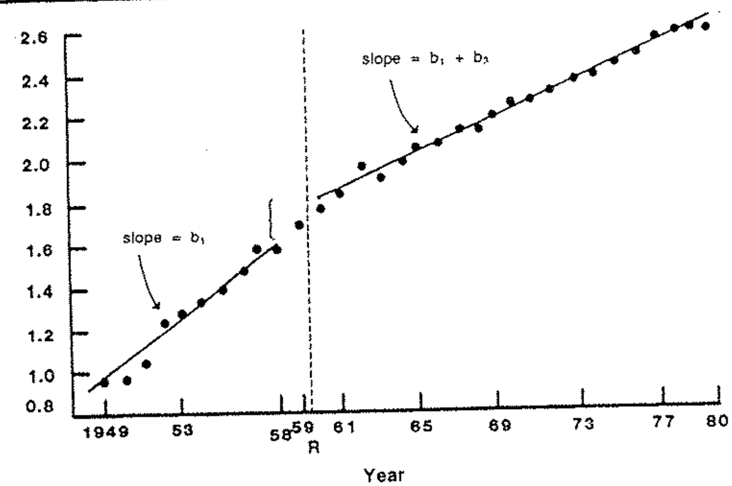
where Y_t = N time-series observations on the dependent variable; X_{1t} = a dummy variable counter for time from 1 to N; X_{2t} = a dichotomous dummy variable scored 0 for observations before the event and 1 for observations after (formerly D_{1t}); X_{3t} = a dummy variable counter of time scored 0 for observations before the event and 1, 2, 3 . . . for observations after the event (formerly D_{2t}). In this model, b_2 estimates the postintervention change in intercept, and b_3 estimates the postintervention change in slope. Perhaps a data example will clarify how this is so.

An earlier investigation used SITS to explore the impact of the Cuban revolution on national economic growth (see Lewis-Beck, 1979, 1981). In the scatterplot of Figure 9.3 are annual observations (1949-1980) on national energy consumption (in millions of metric tons logged to the base e), the proxy variable used for GNP. The vertical line that "interrupts" the series marks the beginning of the Cuban revolution on January 8, 1959, when Castro assumed control. Thus virtually all energy consumption for 1959 and after was under Castro's government, but energy consumption for 1958 and before took place under prerevolutionary regimes. The research question, of course, is whether or not energy consumption changed after the Castro takeover. Visual inspection of the plot hints that it did. In the short run, energy consumption appears to have taken an unusual jump from 1958 to 1959, suggesting an intercept change. In the long run, the trend looks a bit less steep after 1958, implying a slope change. Are these changes apparent rather than real, being in fact no more than products of chance? Further, if they are real changes, exactly how much do they amount to? OLS estimates of the general SITS model provide answers to these queries:

$$\begin{array}{lllll} G_t = .840 & + .076X_{1t} & + .110X_{2t} & - .034X_{3t} & + e_t \\ \quad (34.97) & (19.60) & (4.27) & (-8.43) & \end{array} \qquad [9.10]$$

$$\text{R-squared} = .996 \quad N = 32 \quad \text{D-W} = 1.42$$

where G_t = annual Cuban energy consumption (millions of metric tons, natural logs); X_{1t} = a dummy variable counter for time from 1 to 32; X_{2t} = a dichotomous dummy variable scored 0 for observations before 1959 and 1 for observation 1959 and after; X_{3t} = a dummy vari-



NOTE: R = revolution.

Figure 9.3: Simple Interrupted Time Series Plots of Annual Cuban Energy Consumption (1949-1980)

ables counter for time scored 0 for observations before 1959 and 1 (at 1959), 2 (at 1960), . . . , 22 (at 1980); e_t = the error term; the figures in parentheses = the t-ratios; the R-squared = the coefficient of multiple determination; D-W = the Durbin-Watson statistic; N = the number of annual observations (from 1949-1980, gathered from various issues of the *United Nations Statistical Yearbook*).

According to the t-ratios for b_2 and b_3, respectively, the level and the trend in energy consumption shifted significantly after Castro came to power (for statistical significance at .05, one-tail, $/t/ > 2.05$). Let us consider precisely how the SITS design captures these shifts, initially by examining the trend change. Using equation 9.10, we will predict the energy consumption values first for each of the preintervention years, then for each of the postintervention years. If the predicted values are now plotted, as in Figure 9.3, they are seen to form two straight lines. (These lines would also be produced by the bivariate regression of energy consumption on time, within each period; skeptics should demonstrate this for themselves.) Visually, the slopes of these lines appear different. Indeed, the slope of the preintervention line is .076 (b_1), but the slope of the postintervention line is only .042 ($b_1 - b_3$). As noted, this drop in slope of .034 (b_3) is statistically significant, indicating that the growth of energy consumption has slowed under Castro.

Is N too small?

With regard to the positive change in the level, this is suggested by the gap between the two lines occurring from 1958 to 1959. For the years before the revolution, the prediction equation simplifies to $\hat{G}_t = b_0 + b_1 X_{1t}$ (given that X_{2t} and X_{3t} assume zero values). If the revolution had not taken place, then the prediction for 1959 on the basis of this equation for prerevolutionary Cuba would be 1.676 (that is, $\hat{G}_{59} = .840 + .076(11) = 1.676$). But, the revolution did occur, so the preferred prediction for 1959 becomes 1.752 ($\hat{G}_{59} = .840 + .076(11) + .11(1) - .034(1) = 1.676 + .076 = 1.752$). This .076 increase in the prediction is due to the post-intervention intercept rise (b_2) and would be still larger if the depressing impact of the negative slope change (b_3) were not working, $b_2 - b_3 = .110 - .034 = .076$. (It is strictly coincidental that this value is identical to b_1; in general, this would not be so.)

It is possible to refine further the meaning of these changes. Because the dependent variable is logged, a percentage change interpretation of the parameter estimates can be made (Tufte, 1974: 124–126). For example, a one-year change in X_{1t} produces a (b_1 x 100) percentage change in energy consumption. That is, b_1 estimates that energy consumption increased at the rate of 7.6% a year prior to the revolution. Then, after Castro came to power, the rate dropped 3.4 percentage points, according to b_3. Put another way, the estimated annual rate of energy consumption increase under the revolutionary government is 4.2% ($b_1 - b_3$). Therefore, in the long run, the Cuban revolution is associated with a diminished upward trend in energy consumption. But, in the short run, the picture is different. The estimated growth from 1958 to 1959 is 15.2 %, which reflects the substantial intercept shift measured by $b_2(\hat{G}_{59} - \hat{G}_{58} = b_1 + b_2 + b_3 = .076 + .11 - .034 = .152)$. In sum, this SITS analysis suggests that the arrival of the Cuban revolution was followed in the short term by increased economic growth (indicated by a significant intercept change, b_2), and in the long term by decreased economic growth (indicated by the significant slope change, b_3).

SITS analysis focuses on the impact of a single intervention. Obviously, however, a researcher might also be interested in the impact of multiple interventions. Suppose, to resume the traffic fatalities example, that the Kentucky speed limit was again changed, and this time it was raised to 65 mph. There are now two "interruptions" in the time series of fatalities, which is plotted in Figure 9.2d. These apparent slope and intercept changes can be estimated in a multiple interrupted times series (MITS) model, which straightforwardly extends the SITS model, as follows:

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + b_3 X_{3t} + b_4 X_{4t} + b_5 X_{5t} + e_t \quad [9.11]$$

where Y_t, X_{1t}, X_{2t}, and X_{3t} are defined as in the SITS model of equation 9.9, thereby capturing the first event; X_{4t} = a dichotomous dummy variable scored 0 for observations before the second event and 1 for observations after; X_{5t} = a dummy counter of time scored 0 for observations before the second event and 1, 2, 3 . . . for after the second event; e_t = the error term. As is readily seen, the model is general in form, for any number of interventions can be incorporated by merely adding the appropriate variables.

Although examples of SITS analysis are not uncommon in the research literature, examples of MITS analysis are almost nonexistent, despite its undoubted applicability to numerous research questions. In earlier work, I utilized MITS analysis to assess the effect of three coal mine safety laws (from 1941, 1952, and 1969) on the fatality rate among miners (for details and follow-up studies, see Lewis-Beck and Alford, 1980; Perry, 1982; Nelson and Neumann, 1982). Here, I update these findings to 1980, and use the results to explicate a MITS analysis. Figure 9.4 is a plot of annual observations on the fatality rate in U.S. coal mines from 1932-1980. The vertical lines mark the introduction of the three different pieces of safety legislation. Eyeballing the scatter, it seems that the 1941 and 1969 laws influenced the fatality trend. Further, it appears that the 1941 law did not influence the level, but the 1969 law may have. The other possibilities are difficult to assess visually. To clarify what is going on, I formulate a full MITS model and estimate it with OLS, yielding

$$\begin{aligned} F_t = {} & 1.493 - .001X_{1t} - .096X_{2t} - .045X_{3t} + .151X_{4t} \\ & (18.91) \quad (-.05) \quad (-.95) \quad (-2.66) \quad (1.72) \\ & .047X_{5t} - .228X_{6t} - .058X_{7t} + e_t \qquad [9.12] \\ & (3.76) \quad (-2.48) \quad (-4.71) \\ & \text{R-squared} = .916 \quad N = 49 \quad \text{D-W} = 2.28 \end{aligned}$$

where F_t = annual fatalities per million hours worked; X_{1t} = a dummy counter for years, from 1 to 49; X_{2t} = a dichotomous dummy variable scored 0 for observations before 1942 and 1 for 1942 and after; X_{3t} = a dummy counter for years, scored 0 for observations before 1942, and 1, 2, 3 . . . . for 1942 and after; X_{4t} = a dichotomous dummy variable scored 0 for observations before 1953 and 1 for 1953 and

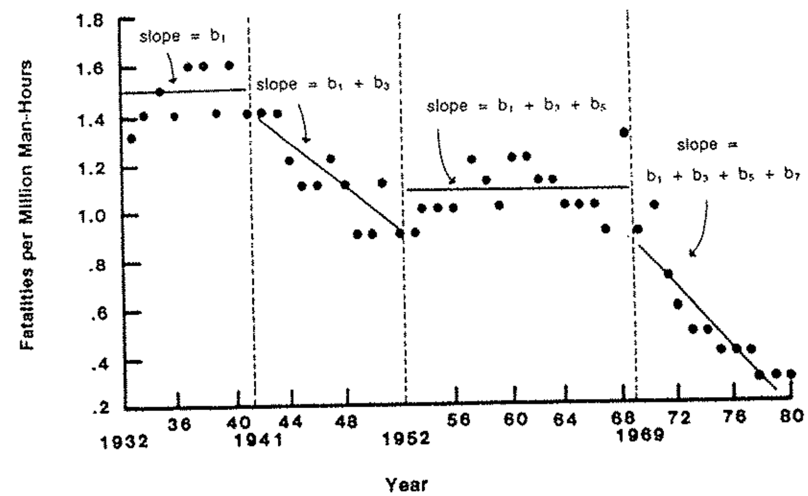


Figure 9.4: Multiple Interrupted Time Series Plots of the Annual Coal Mining Fatality Rate (1932-1980)

after; X_{5t} = a dummy counter for years, scored 0 for observations before 1953 and 1, 2, 3 . . . . for 1953 and after; X_{6t} = a dichotomous dummy variable scored 0 for observations before 1970 and 1 for 1970 and after; X_{7t} = a counter for years scored 0 for observations before 1970 and 1, 2, 3 . . . . for 1970 and after; e_t = the error term; the figures in parentheses = the t-ratios; the R-squared = the coefficient of multiple determination; N = number of annual observations (1932-1980); D-W = the Durbin-Watson statistic; the data are gathered from various publications of the *Bureau of Mines* and its successors.

The parameter estimates $\hat{b}_0$ and $\hat{b}_1$ indicate, respectively, the level and trend of the time series before imposition of the 1941 law. In order to assess whether b_0 and b_1 were changed by this new legislation, it is necessary to study b_2 and b_3. Given that, according to the t-ratios, the estimate for b_2 is not significantly different from zero, the inference is that these regulations had no impact on the level of fatalities (for statistical significance at .05, two-tail, $/t/ > 2.2$). However, the b_3 estimate is significant, which indicates that the 1941 law did change the slope of the time series, substantially lowering the fatality rate. Indeed, the slope before the 1941 legislation is $\hat{b}_1$ (−.001), whereas after the 1941 legislation it is $\hat{b}_1 + \hat{b}_3$ = −.046, showing that the fatality rate changed from a stable to a declining condition.

The 1952 and 1969 interventions are evaluated in the same manner. One observes that the 1952 legislation has an insignificant intercept effect (see $\hat{b}_4$), but a significant positive slope effect (see $\hat{b}_5$). The result is that the 1952 legislation apparently arrested the downward movement in the fatality rate instituted by the 1941 legislation. (The trend line for this post-1952 period is flat, $\hat{b}_1 + \hat{b}_3 + \hat{b}_5 = -.001 - .045 + .047 \cong 0$). Finally, with respect to the 1969 regulations, the estimates for b_6 and b_7 suggest they brought about significant reductions in the fatality rate, in both the short and the long run. The slope of the time series in the post-1969 period equals $b_1 + b_3 + b_5 + b_7$. In general, the slope of the time series within any postintervention period, for example, post-1969, can be estimated by adding to the initial slope estimate (for example, $\hat{b}_1$) all the slope change parameter estimates calculated up to and including that period, e.g., $\hat{b}_3 + \hat{b}_5 + \hat{b}_7$. (In Figure 9.4, observe that if one fits a simple regression line within each time period, each has a slope equal to this combination of MITS coefficients.) Overall, this collection of dummy variables, which have been scored to reflect multiple federal interventions in the area of coal mine safety, does an excellent job of accounting for the changes in the mining fatality rate, as the R-squared = .92 demonstrates.

PROBLEMS OF INFERENCE

From an ITS analysis, one hopes to make causal inferences. Did the intervention produce the change observed? Fortunately, the ITS design, which is quasi-experimental, seems relatively strong in this regard. One can observe the pattern of behavior in Y before the intervention occurs, then after. In appearance, this is similar to an experiment with a control group and a treatment group. Experimental design, of course, allows the surest inferences about causality. However, what is missing here with an ITS analysis is random assignment to groups, which means it remains an essentially nonexperimental approach. Therefore, in order to make causal inferences, it is necessary to meet the assumptions of a nonexperimental multiple regression model. These assumptions include absence of specification error for the model, absence of measurement error in the variables, absence of perfect multicollinearity among the independent variables and—for the error term—a zero mean, homoscedasticity, no autocorrelation, no correlation with independent variables, and normality. (For regression analysis generally, these assumptions are discussed elsewhere; Berry and Feldman, 1985; Lewis-Beck, 1980.) Below, I give particular attention to specification error, measurement error, and perfect multicollinearity. Lastly, I consider the no-autocorrelation assumption, too often the exclusive focus of concern in other treatments of ITS analysis.

THE PROBLEM OF SPECIFICATION ERROR

A regression model is properly specified when each independent variable is linearly related to the dependent variable, no irrelevant independent variables have been included, and no relevant independent variables have been excluded. Let me address how each of these conditions apply in an ITS analysis. With respect to linearity, the best equation in the pre- and postintervention periods must yield a straight line. One observes this most clearly in Figure 9.3 of the Cuban example, where the trends are straight both before and after Castro took control. Overall, the linearity assumption appears quite sound for this SITS model which, it should be recalled, produced an R-squared = .996, an almost perfect linear fit. Obviously, however, a time trend need not always be straight, as it is not in the diagram of Figure 9.5a, in which Y follows an exponential path. If an ordinary SITS model were estimated with these data, it would fit two straight lines (see broken lines) to the curved trends on either side of the intervention. This erroneous specification leads to faulty inference from these ITS estimates, which would indicate a large positive slope shift ($+b_3$).

The solution for this predicament is a transformation that will make the relationship linear, thus allowing the proper use of OLS. The pattern in Figure 9.5a is a fairly common one, depicting the trajectory of an absolute growth figure, such as GNP, over time. Such trends can usually be linearized with a logarithmic transformation of the dependent variable. Indeed, recall this was initially done with the data in the Cuban revolution example, where the energy consumption variable was logged in order to overcome such curvilinearity. In general, as with any regression analysis, if the scatterplot of the raw observations reveals nonlinearity, a transformation should be sought to make the relationship linear. A variety of variable transformations—logarithmic, reciprocal, square root, polynomial—can be used (see Berry and Feldman, 1985, chap. 5). Careful choice here is important, for an improper transformation could increase, rather than decrease, the observed nonlinearity. The preferred transformation depends heavily, of course, upon the shape of the nonlinearity. Look, for example, at the nature of the nonlinearity in the postintervention period in Figure 9.5b. Because of the inverted bowl shape, a second-order polynomial is suggested, which would involve adding a variable,

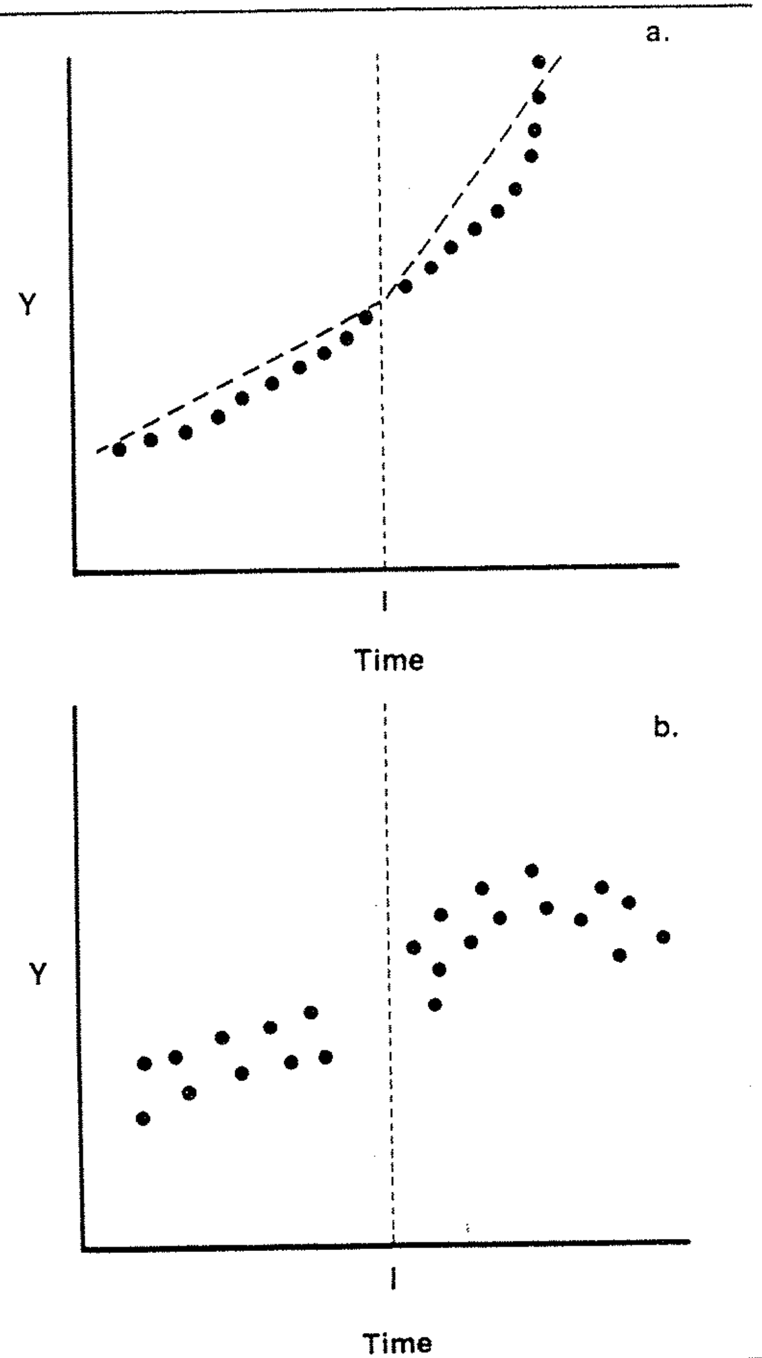


Figure 9.5 a, b: Examples of Nonlinear Interrupted Time Series

X^2_{3t}, the square of the postintervention time counter, to the original SITS equation. If this variable, X^2_{3t}, had a significant coefficient, one might infer that the intervention had changed the shape of trend. But, it may not be clear, from visual inspection alone, which transformation should be chosen. For instance, in Figure 9.5b, the log or reciprocal of X_{3t} might also be candidates because of the decelerating increase visible in the time series. In this case, the rival transformations could be performed, and the changes in the R-squared from one model to the next evaluated. Overall, if no one transformation seems superior to another (or to the untransformed model) in improving the goodness of fit, then the choice of models should be based on other criteria, such as interpretability of the results or preservation of degrees of freedom.

Let us now consider the specification error of including irrelevant variables. For an ITS model, this means including a dummy variable to capture a nonexistent intercept or slope change. The standard method for assessing whether or not such an error has been committed is the application of a significance test. If the parameter estimate for the variable fails to achieve statistical significance, then the conclusion is that the variable should be omitted from the equation. (Of course, this decision would be wrong one time in twenty, assuming a .05 significance level.) Using the rule for the coal mining fatalities equation, it appears that the intercept change variables X_{2t} and X_{4t} should be excluded in a reestimation, which I carry out below.

The last aspect of specification error, the exclusion of relevant variables, raises the question of spuriousness. An estimated ITS equation may indicate that an intervention significantly alters the phenomenon under study. However, it is possible that this impact is only apparent, the product of the common prior influence of a "third variable" (Z_t) on the intervention and on the dependent variable. Suppose, in the coal mine example, that fatalities are more likely to occur in underground mines, measured by Z_{1t}, and in small mines, measured by Z_{2t}; further, suppose that when more small, underground mines are in operation, the federal government more vigorously intervenes with safety regulations. In this situation, the significant coefficients of the MITS safety regulation variables might be spurious, having been brought about by the joint impact of these "third variables," Z_{1t} and Z_{2t}. To test for this possibility, these variables were introduced into the more properly specified MITS equation (which excludes the insignificant X_{2t} and X_{4t}), yielding the following OLS estimates:

$$\underset{}{F_t} = \underset{(.06)}{.030} - \underset{(-.59)}{.006X_{1t}} - \underset{(-2.89)}{.042X_{3t}} + \underset{(5.32)}{.057X_{5t}} - \underset{(-1.36)}{.128X_{6t}}$$

$$-.050X_{7t} + .87Z_{1t} + .010Z_{2t} + e_t \qquad [9.13]$$
$$(-4.00) \quad (1.38) \quad (2.08)$$
$$\text{R-squared} = .928 \quad N = 49 \quad \text{D-W} = 2.33$$

where F_t, X_{1t}, X_{3t}, X_{5t}, X_{6t}, X_{7t} and the statistics are defined as with equation 9.12; Z_{1t} = annual percentage of miners working underground; Z_{2t} = annual percentage of mines producing less than 50,000 tons (data on Z_{1t} and Z_{2t} are from *Minerals Yearbook*, various issues).

According to these results, the significant, long-run slope changes ($\hat{b}_3$, $\hat{b}_5$, $\hat{b}_7$) captured in the initial MITS model (equation 9.12) are not spurious. But, the introduction of controls, Z_{1t} and Z_{2t}, indicates that the significant short-run intercept change (b_6) of the initial model was spurious. Moreover, the findings of equation 9.13 suggest, quite plausibly, that the fatality rate is responsive to something besides federal regulation, namely, mine size. (However, the other control variable, percentage of miners underground, appears not to affect fatalities.) These pieces of evidence enable further improvements in model specification. Coal mining fatalities seem to be a function of the long-run impact of federal safety legislation in 1941, 1952, and 1969 (X_{3t}, X_{5t}, X_{7t}), plus mine size (Z_{2t}). Estimating such a revised model with OLS yields

$$F_t = .315 - .014X_{1t} - .039X_{3t} + .060X_{5t} - .063X_{7t} + .017Z_{2t} + e_t$$
$$(1.05) \quad (-1.43) \quad (-2.62) \quad (5.98) \quad (-5.76) \quad (4.21)$$
$$\text{R-squared} = .921 \quad N = 49 \quad \text{D-W} = 2.10 \qquad [9.14]$$

where the definitions of the variables and the statistics are as with equation 9.13. This final specification is an advance over the initial specification in several important ways: (1) it is more parsimonious, explaining the same amount of variance with fewer independent variables; (2) it is theoretically more plausible, indicating that federal safety legislation yields long-term rather than short-term results; and, (3) it is theoretically more complete, incorporating a substantive variable, that is, coal mine size, into the model along with the dummy variable time counters.

With ITS analysis, there is also a particular kind of specification error involving the omission of a relevant variable, which comes from the occasional seasonality of time-series data. Although relatively rare in annual series, cycles may be exhibited in quarterly or monthly series. The classic example is the December jump in department store sales. A dummy variable would easily incorporate such a twelfth-month rise into an ITS model. Suppose, for instance, that our Kentucky traffic fatalities data

were recorded on a monthly, rather than an annual, basis. If we repeatedly observed a leap in fatalities in December, perhaps due to the holiday season, then we might develop a monthly SITS equation that included a December dummy, scored 1 = December and 0 = other months. Then, the slope and intercept changes produced by the speeding crackdown could be assessed free of any distorting influence from the December cycle.

At this point, I would like to consider how to reduce the risk of specification error by drawing on corroborative evidence from the analysis of other, relevant, time series. One type of relevant time series is a related dependent variable that would be expected to respond similarly to the intervention. For example, in my coal mine safety investigation, I reasoned that actual safety enforcement activity (measured by the annual size of the health and safety budget) should mirror the pattern of interruptions in the fatality rate, if in fact the federal regulations were operating as postulated. One observes that this is actually so, by eyeballing the time-series plot of the coal mine health and safety budget (natural logs) in Figure 9.6 (see Lewis-Beck and Alford, 1980). For instance, the trend changes in the pattern of budget growth from period to period appear to reflect the changes in the law. Estimation of the full MITS model, with an R-squared = .99, nicely confirms this general picture. (In addition, the significant short-term budget spurts would seem plausible, immediate consequences of the introduction of new legislation).

$$B_t = 6.497 - .007X_{1t} + .582X_{2t} + .135X_{3t} - .125X_{4t}$$
$$(58.5) \quad (-.4) \quad (4.1) \quad (5.7) \quad (-1.0)$$
$$-.067X_{5t} + 1.174X_{6t} + .007X_{7t} + e_t \qquad [9.15]$$
$$(-3.9) \quad (7.5) \quad (2.4)$$
$$\text{R-squared} = .99 \quad N = 45 \quad \text{D-W} = 1.5$$

where B_t = annual coal mine health and safety budget (thousands of dollars, natural logs) 1932-1976, (gathered from the *Budget of the United States Government*, various issues); the other variables are defined as in equation 9.12.

Another kind of relevant time series useful for checking specification error is a "nonequivalent no-treatment control group time series"; that is, a time series from a population that is as similar as possible to the one under study, except it has not experienced the intervention. Applying this idea to the Kentucky speeding crackdown example, one might put together the time series of traffic fatalities in Tennessee, which had had no

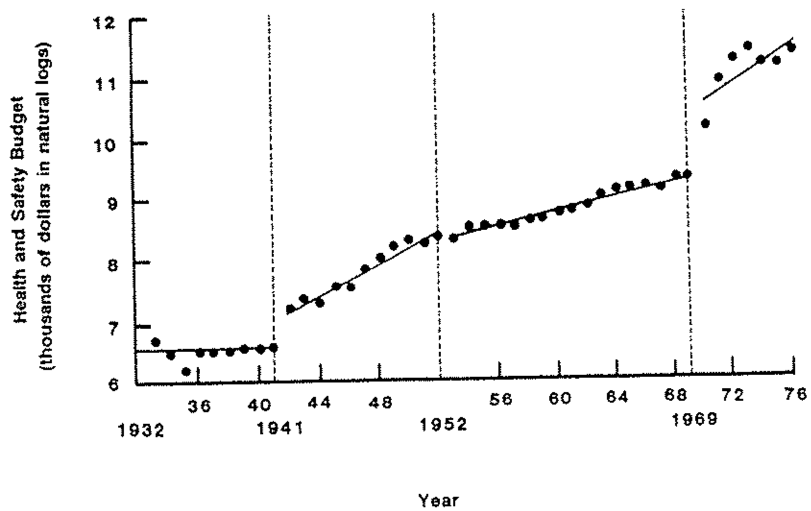


Figure 9.6: MITS Analysis of Federal Coal Mine Health and Safety Budget

crackdown. On the one hand, then, if the Kentucky series manifests a significant slope change and the Tennessee series does not, one gains confidence that the speeding limitation caused the fatality rate reduction. On the other hand, if both states show a significant decline in fatalities after 1972, then the suspicion would be that the Kentucky results were not caused by the new speed limit. Rather, the decline in both states would then appear to be a product of a more general influence affecting both states at the same time, such as (hypothetically) the introduction of a national auto seatbelt requirement in 1972.

PROBLEMS OF MULTICOLLINEARITY

The foregoing example leads to the problem of perfect multicollinearity in an ITS context. The intervention may occur at the same time as another possibly relevant event. Suppose, for example, that the Kentucky traffic safety researchers hypothesize that the 1972 speeding crackdown changed the level of the time series, and therefore they formulate a model with D_t scored 0 before 1972 and 1 for 1972 and after. However, they realize that in the same year the national seatbelt law was introduced; hence they try to control for this rival influence on fatalities by introducing a

third variable, E_t, scored 0 before 1972 and 1 for 1972 and after. Of course, D_t and E_t will be perfectly correlated, estimation will be impossible, and the separate influence of these two interventions will be totally unascertainable. Perfect multicollinearity poses an intractable problem for ITS analysis, as for regression analysis generally. But a potential solution lies in making the multicollinearity less than perfect. Specifically, one may be able to locate a proxy variable for the rival event that is not perfectly correlated with the intervention dummy. For example, in the traffic fatalities example, if the researchers are concerned about controlling for the influence of seatbelt usage, they might introduce into the ITS model the continuous variable, S_t, the annual percentage of car owners regularly employing seatbelts. The correlation between D_t and S_t would certainly not be perfect, and might allow the researchers to assess the impact of the speeding crackdown free of the confounding influence of seatbelt usage.

The substitution of such a proxy variable could still leave a high level of multicollinearity, which might generate a statistically insignificant parameter estimate even though the intervention brought about real change in the population (on the multicollinearity problem, see Berry and Feldman, 1985, chap. 4; Lewis-Beck, 1980: 58–63). Generally, in an ITS model, high multicollinearity may be introduced if one attempts to include a rival event that occurred at almost the same time as the intervention of interest. In the Cuban example, is the revolution responsible for the economic growth pattern, or is the U.S. embargo, which began in 1960, responsible? One approach is to simply think of the embargo as part of the revolutionary experiment, which included a shift in trade partners from the United States to the Soviet Union. Then, the embargo is captured, along with the other aspects of the revolution, in the ITS dummies of the original model (see equation 9.10). However, if one wishes to distinguish conceptually between the domestic and the international aspects of revolutionary change, a solution is more difficult. The U.S. embargo took place a year after the revolution began; therefore, a dummy time counter for the embargo would have the same values as X_{3t} in the original model, except for the 1959 observation. This high degree of multicollinearity makes reliable separation of the slope changes impossible. (However, it does not prohibit assessment of the intercept shift, from 1958 to 1959. For that brief period, there was domestic revolution operating without the potentially confounding effects of the U.S. embargo. As the coefficient for X_{2t} in the original model shows, energy consumption jumped significantly in this period, hinting that, at least in the short run, the revolution positively influenced growth.) Broadly speaking, in ITS analysis, when several interventions are being considered that are fairly closely spaced in time, one should be alert to the possibility of multicollinearity problems. Insignificant coefficients may simply be the product of a high degree of linear dependency among the intervention dummy variables.

Another way in which high multicollinearity can arise with ITS analysis is when the number of observations before and after the event are seriously uneven (recall Figures 9.1d–9.1e). For instance, if the researcher has a much longer string of observations before the event, then X_{2t} and X_{3t} in the SITS model (equation 9.9) will tend to be highly correlated because the large number of pre-event values will equal zero. For example, in my study of the Cuban revolution, I also carried out a SITS analysis of annual sugar production from 1929-1974. The SITS results showed that the coefficients for X_{2t} and X_{3t} were statistically insignificant (Lewis-Beck, 1979). However, I worried that this lack of significance was a product of the high correlation between X_{2t} and X_{3t}, which was .85. (This correlation results from the fact that there were 39 observations before the Castro takeover compared to 16 after, which means that X_{2t} and X_{3t} share many identical, zero, values). I explored the drastic solution of dropping first X_{2t}, then X_{3t}, from the model. (Such a solution is drastic because it implies willful commission of specification error. Technically, the preferred solution is the standard econometric recommendation of increasing the sample size, which is usually not possible in practice.) Thus each variable, in turn, is allowed to account for as much sugar production as it can, independent of the control of the other. Still, neither X_{2t} nor X_{3t} managed to exhibit a significant impact. Therefore, I concluded that there was no multicollinearity problem, and that the insignificant results of the original model were valid.

SOME MEASUREMENT ERROR PROBLEMS

Measurement error problems can take a special form with ITS analysis. In order to construct the independent variables, the intervention must be clearly marked in time. When exactly did the event occur? In the case of the Cuban revolution, Batista fled the island on December 31, 1958, and Castro marched victoriously into Havana on January 8, 1959. Therefore, given that the data are annual observations, I regarded 1959 as the year the revolution began. The next question is, "Should the year (quarter, month) of the event be scored on the "before" or "after" side of the intervention dummies?" With annual data, if the event took place in the first half of the year, then that whole year should generally be coded as a postintervention value, because the behavior of the dependent variable would happen mostly after the event. For example, because Castro came to power in January 1959, X_{2t} and X_{3t} in the basic SITS model are scored "1" for the year 1959 (see equation 9.10).

One problem that can arise is uncertainty as to the precise date the event occurred. Although the ITS model requires definite cut-points, we know that social processes are not always so easily pulled apart. For instance, suppose someone argued that the Cuban revolution did not really get underway until 1960. If this is so, then year 1 for the ITS dummies should be 1960, rather than 1959. One strategy for resolving this disagreement is to specify different intervention points, estimate the rival ITS models, and observe which more clearly supports the dominant hypothesis. With the Cuban example, I altered the intervention point to 1960, 1961, 1962, and 1963, respectively. (That is, I reestimated the SITS model, each time making these years the first postintervention observation, with a score of 1 on X_{2t} and X_{3t}.) These estimates confirmed my preference for the original intervention scoring (1959 = 1). In particular, the various estimates for b_2 steadily decline as the intervention point is moved away from 1959, and actually cease to be significant at 1962 = 1. These results clearly support the impression of careful observers of the Cuban scene that 1959 was a revolutionary peak in economic growth.

The last difficulty that crops up is whether or not the intervention effect occurs with a lag. We may expect the effect of the intervention to appear after the passage of some time, rather than immediately. In our hypothetical example of the 1972 Kentucky speeding crackdown, for example, we might think that the law would not significantly affect fatalities until the following year, because it takes people a while to learn of the law and see that it is actually being enforced. If the law is operating with this one-year lag, then the preferred scoring of the SITS dummy variables X_{2t} and X_{3t} would be year 1 = 1973 (instead of 1972).

THE PROBLEM OF AUTOCORRELATION

Finally, I turn to the regression assumption of no autocorrelation. In order to make valid causal inferences from the OLS estimates of an ITS equation, error terms must not be correlated, that is, $E(e_t\ e_{t-j}) = 0$, where $j \neq 0$ (for a general reference on this problem in time series, see Ostrom, 1978). This assumption is routinely violated in the analysis of time-series data (data involving repeated observations on the same unit of analysis over time). Almost always, the autocorrelation is positive. In the presence of positive autocorrelation, the regression residuals ($\hat{e}_t = Y_t - \hat{Y}_t$) will tend to track each other when plotted across time, with positive residuals followed by positive residuals until a "shock" occurs, then negative residuals follow negative residuals. It is not difficult to imagine why such a pattern appears so often with time series. The error term of the regression equation stands for an ensemble of variables that influence Y but are left out of the model. These variables shaping Y at time t-1 would likely be related to the variables shaping Y at time t, which implies autocorrelation. Consider the following simple model, in which the annual United States defense budget (Y_t) is a function of the annual presidential defense budget request (X_t):

$$Y_t = b_0 + b_1X_t + e_t \qquad [9.16]$$

Upon contemplation of what variables are likely in that collection represented by the error term, e_t, a candidate that comes easily to mind is the annual Gross National Product (GNP). Obviously, GNP from the prior year, GNP_{t-1}, is correlated with GNP from the present year, GNP_t; there-fore, the correlation between e_t and e_{t-1} is not equal to zero. Such a pattern, in which error from the immediately prior time (e_{t-1}) is correlated with error at the current time (e_t), depicts a first-order autoregressive process, AR(1), by far the most common form.

In the face of this autocorrelation, the OLS parameter estimates, $\hat{b}_0 + \hat{b}_1$, nevertheless remain unbiased. Hence when the primary focus of the ITS analysis is on the magnitude of the parameter changes brought about by an intervention unambiguously known to have altered the level and/or slope of the series, then the autocorrelation problem is less impor-tant. However, such certain knowledge about an intervention's effective-ness is usually lacking. For most ITS analysis, the critical question is not, How big was the effect of the intervention?; rather, it is simply, Did the intervention have an effect? When the latter question is paramount, the presence of autocorrelation is problematic for OLS estimation because it invalidates the significance tests, our essential tool for assessing whether the intervention had an effect.

These significance tests are rendered invalid due to the bias that autocorrelation creates in the estimation of the coefficient variances. These estimated variances are typically too small, given that most autocorrelation is positive. The more the true variance is underestimated, the more the t-ratio from the OLS results is inflated, as its formula makes clear:

$$t = \frac{\hat{b}}{\sqrt{\text{var } \hat{b}}} \qquad [9.17]$$

The underestimated denominator gives the t-ratio too large a value. It can too easily surpass the value of 2.00, thus erroneously suggesting significance at the .05 level, with a two-tailed test. Thus, in the presence of positive autocorrelation, OLS tends to make the coefficients appear significant when they are not. We risk, in other words, committing the Type I error of rejecting the null hypothesis when it is correct. Further, this risk is far from trivial, for this inflation of the t-ratio can be substantial. Indeed, according to McDowall and his colleagues (1980: 13), it is commonly inflated 300% to 400%. The implication is that, in the face of autocorrelation, one cannot safely infer significance from OLS t-ratios even when they well exceed the rule-of-thumb value of 2.00.

Clearly, any ITS analysis that aims to assess the statistical significance of an intervention must explicitly deal with the autocorrelation problem. Until the data suggest otherwise, the presumption must be that the error terms of an OLS model are not independent. There are two basic types of dependent error processes, the autoregressive and the moving average. A positive first-order autoregressive process, AR(1), by far the most common, takes the form

$$e_t = \rho e_{t-1} + u_t \qquad [9.18]$$

where e_t = the error term; ρ = rho, the population correlation between e_t and e_{t-1}. Here, the error at time t is a function of the error at time t-1, plus a random variable, u_t. An AR(1) model yields a particular "autocorrelation function" such as that in Figure 9.7a, which shows the correlation of the error terms at an increasing lag. This theoretical correlogram graphs the correlation of $e_t\, e_{t-1}$ (A_1), $e_t\, e_{t-2}$ (A_2), $e_t\, e_{t-3}$ (A_3) . . . and $e_t\, e_{t-10}$ (A_{10}). As with all positive AR(1) error processes, these correlations follow a steady exponential decline. Very rarely, a first-order autoregressive process is negative,

AR(1)

$$e_t = -\rho e_{t-1} + u_t \qquad [9.19]$$

In this case, the theoretical correlogram looks like Figure 9.7b, where the autocorrelation functions oscillate from positive to negative as they exponentially decline.

The moving-average process is the second basic error pattern. A positive first-order moving-average process, MA(1), takes the form,

MA(1)

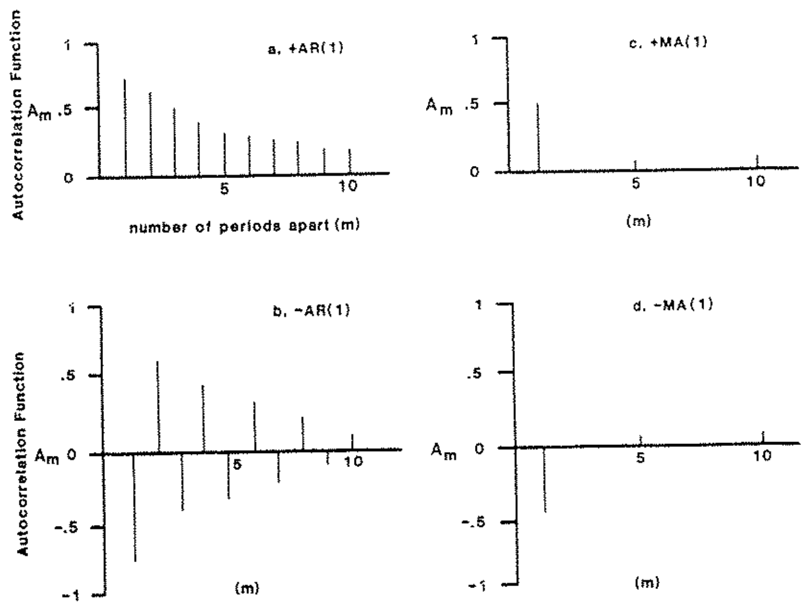


Figure 9.7 a-d: Theoretical Autocorrelation Functions ($A_m = \rho_{e_t e_{t-m}}$) for Different Error Processes

$$e_t = u_t + \rho u_{t-1} \qquad [9.20]$$

where the error at time t is strictly a function of a random variable at time t, plus a portion of that random variable at time t-1. The autocorrelation function of a moving average process is quite different from an autoregressive process. The theoretical correlograms in Figures 9.7c and 9.7d describe, respectively, a positive and a negative MA(1) pattern. The distinguishing characteristic of an MA(1) error process is a spike at the first lag, followed by correlations of zero at greater lags.

In practice, one may encounter a mixed autoregressive-moving average process. Further, higher-order processes are possible. For example, an AR(2) model takes the form,

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t \qquad [9.21]$$

However, mixed processes and higher-order processes are in fact very rare in time series of social science data (McDowall et al., 1980: 28-29, 46). Instead, evidence indicates that positive AR(1) processes dominate

these series. For instance, Ames and Reiter (1961) examine 100 annual time series randomly selected from statistical abstracts and report that, on average, the autocorrelation function exhibits exponential decay. Thus because any error process of error dependency is almost always AR(1), the discussion below focuses on its detection and correction. (Other forms of error are generally less tractable; for an introduction to estimation procedures in this case, see Hibbs, 1974.)

The most widely used method of assessing the presence of first-order autocorrelation is the Durbin-Watson statistic, which has the following formula:

$$\text{D-W} = \frac{\sum_{t=2}^{N} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{N} \hat{e}_t^2} \qquad [9.22]$$

where $\hat{e}_t$ = OLS residuals. Glancing at this formula, one sees that the more positive the autocorrelation the smaller the numerator, thus diminishing the D-W value. In opposite fashion, the more negative the autocorrelation, the larger the D-W value. More formally, assuming first-order autocorrelation, the following generalizations hold for ρ (rho):

$$\text{if } \rho = +1 \quad \text{then D-W} \cong 0$$
$$\text{if } \rho = -1 \quad \text{then D-W} \cong 4$$
$$\text{if } \rho = 0 \quad \text{then D-W} \cong 2.$$

Hence, in terms of accepting the null hypothesis of no autocorrelation, the closer D-W is to 2.00, the better. To evaluate precisely whether a particular D-W value allows acceptance of the null, one consults a D-W table (see any standard econometrics text). Assuming the table provides a .05, two-tail significance test, we look up the values appropriate to the number of observations, N, and the number of independent variables, K. Two values will be provided, a lower-bound, d_l, and an upper-bound, d_u. With positive autocorrelation, a D-W value less than d_l leads to rejection of the null hypothesis of no autocorrelation, whereas a D-W value exceeding d_u leads to acceptance of the null of no autocorrelation. Unfortunately, if the D-W value falls between d_l and d_u, then there is an uncertainty over whether to accept or reject the null hypothesis. (This uncertainty exists because, in this region, apparent autocorrelation of the error terms may actually be due to autocorrelation of the independent variables; Pindyck and Rubinfeld, 1976: 115.) The major drawback, then, of the D-W test for first-order autocorrelation is the presence of this range of uncertainty, which may prevent a clear decision. Therefore, one might prefer a less ambiguous statistic. Rao and Griliches (1969) conclude, from Monte Carlo experiments, that when an AR(1) error process with $p \geq .30$ exists, OLS should not be employed. With this rule of thumb, we would simply decide significant first-order autocorrelation was present when the estimate of p surpassed .30. (The estimate of p comes from correlating the residuals for the original OLS equation, $\hat{e}_t$ and $\hat{e}_{t-1}$. Note that, because of the lag, one case would necessarily be lost.)

If the ITS analyst, upon application of an appropriate test, discovers a significant AR(1) process, then the OLS t-ratios can no longer be safely employed in making causal inferences. ~~A straightforward remedy is the~~ method of generalized differences, which involves the transformation of the variables using the autocorrelation coefficient, p, thereby rendering the error terms independent. With this assumption met, it is then correct to apply OLS.

Suppose a very simple ITS model were used positing only a slope change, the results of which are generalizable to more complicated cases,

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + e_t \qquad [9.23]$$

where Y_t = N time series observations on the dependent variable; X_{1t} = a dummy time counter from 1 to N; X_{2t} = a dummy variable scored 0 before the event and 1, 2, 3 . . . for observations after the event; e_t = the error term. Assume, further, that the following first-order autoregressive process is operating,

$$e_t = p\,e_{t-1} + u_t \qquad [9.24]$$

Applying generalized differences to these variables yields the following transformed equation:

$$Y_t^* = b_0^* + b_1X_{1t}^* + b_2X_{2t}^* + e_t^* \qquad [9.25]$$

where

$$Y_t^* = Y_t - pY_{t-1}$$

$$X_{1t}^* = X_{1t} - pX_{1t-1}$$

$$X_{2t}^* = X_{2t} - pX_{2t-1}$$

$$b_0^* = b_0(1 - p)$$

$$e_t^* = e_t - pe_{t-1} = (pe_{t-1} + u_t) - pe_{t-1} = u_t$$

Because the revised error term, e_t^*, is now equal to the random variable, u_t, the regression assumption of independent error terms is met, and OLS can be appropriately applied to the transformed variables in order to secure efficient parameter estimates (see Hibbs, 1974: 269; Pindyck and Rubinfeld, 1976: 108-110). In practice, of course, we seldom actually know the autocorrelation coefficient, p, and so we must utilize an estimate, $\hat{p}$.

In sum, a generalized difference approach to correcting for first-order autoregression includes estimation of the autocorrelation coefficient, transformation of the variables with that estimated autocorrelation coefficient, and application of OLS to the transformed variables. Below, I describe the steps of such a procedure. For simplicity, I develop the SITS model shown above.

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + e_t \qquad [9.26]$$

Step 1. Estimate the ITS equation with OLS.

Step 2. Calculate the residuals, $\hat{e}_{1t}, \hat{e}_{2t} \ldots \hat{e}_{nt}$.

Step 3. Obtain the first-order autocorrelation estimate, $\hat{p}$, by correlating $\hat{e}_t$ with $\hat{e}_{t-1}$.

Step 4. Construct the transformed variables, $Y_t^* = Y_t - \hat{p}Y_{t-1}$, $X_{1t}^* = X_{1t} - \hat{p}X_{1t-1}$, $X_{2t}^* = X_{2t} - \hat{p}X_{2t-1}$. (Note that this lagging necessarily involves the loss of one observation from the beginning of the time period.)

Step 5. Get efficient parameter estimates by applying OLS to the following transformed equation:

$$Y_t^* = b_0^* + b_1X_{1t}^* + b_2X_{2t}^* + e_t^* \qquad [9.27]$$

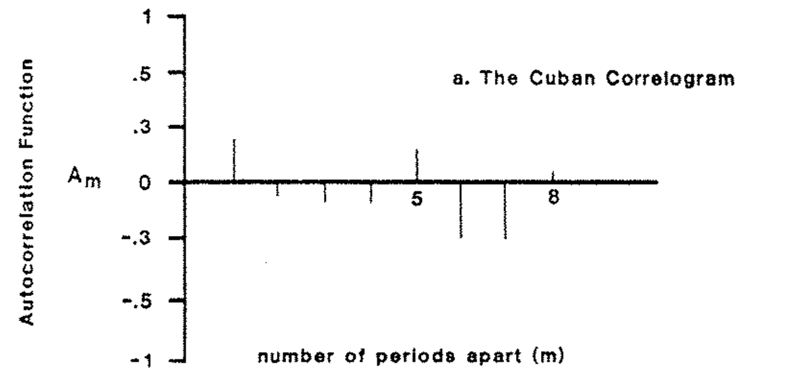
The parameter estimates from this transformed equation have the desirable properties of consistency and efficiency. Further, the resulting t-ratios provide a more valid assessment of the slope change postulated in this model. (Strictly speaking, these t-ratios are approximations in finite samples; see Kelejian and Oates, 1974: 200.) This is a "two-stage" technique, so called because it estimates the parameters only twice (at Step 1 and at Step 5). It is possible to employ an iterative technique, such as the

Cochrane-Orcutt procedure, involving a repetition of the above described steps until the parameter estimates stabilize, which in my experience typically occurs after a few rounds. However, the two-stage procedure described above generates estimators with properties identical to those from Cochrane-Orcutt, at least for large samples. Furthermore, this two-stage procedure has the advantage of being easily computed without a special computer program. (On these matters, see Kmenta, 1971: 287-289; Kelejian and Oates, 1974: 200; Pindyck and Rubinfeld, 1976: 111-112; Ostrom, 1978: 39-40.)

Let us now apply this general discussion of autocorrelation to the specific ITS data examples at hand, the one from the Cuban revolution and the other from U.S. coal mine safety. With regard to the latter, an initial assessment of the presence of first-order autocorrelation comes from a glance at the D-W statistic. In the final specification for the coal mining fatalities equation (that is, equation 14), the D-W = 2.10 which, by its closeness to the rule-of-thumb value of 2.00, indicates that first-order autocorrelation is quite unlikely. More precisely, given a two-tailed, .05 significance test, with 49 observations and 5 independent variables, the D–W value must fall between 1.69 (d_u) and 2.31 (4-d_u) so that we can accept the null hypothesis. Because this D–W value of 2.10 clearly lies within this range, we accept the null hypothesis, concluding that no significant first-order autocorrelation is operating.

Turning to the SITS equation on Cuban energy consumption (equation 9.10), the picture is more cloudy. Here, the D–W statistic is 1.42, which lies in an uncertainty region. (With a two-tailed, .05 significance test, 32 observations, and 3 independent variables, $d_L = 1.16$ and $d_u = 1.55$). Thus on the basis of this test alone, it is not clear whether or not significant autocorrelation exists. This is a case in which the Rao and Griliches rule of thumb for diagnosing a first-order autocorrelation problem, that is, $p < .30$, is useful. Calculating the correlation, $\acute{p}$, between the OLS residuals, $\acute{e}_t$ and $\acute{e}_{t-1}$, yields $\acute{p} = .25$. By this criterion, then, OLS still appears preferred. Therefore, I resolve my uncertainty in favor of the null hypothesis of no first-order autocorrelation.

Despite these assurances, the possibility remains that some other error dependency process is working, besides AR(1). Because different error processes generate different correlograms, it is helpful to calculate the autocorrelation functions for these examples. In composing an empirical correlogram, a decision must be made regarding the length of the last lag, for effective sample size quickly dwindles, increasing the imprecision of the estimates. As a rule of thumb, it has been recommended that these autocorrelations be calculated to lag N/4 (see Hibbs, 1974, p. 280). For the coal mine example, this means computing about 12 empirical



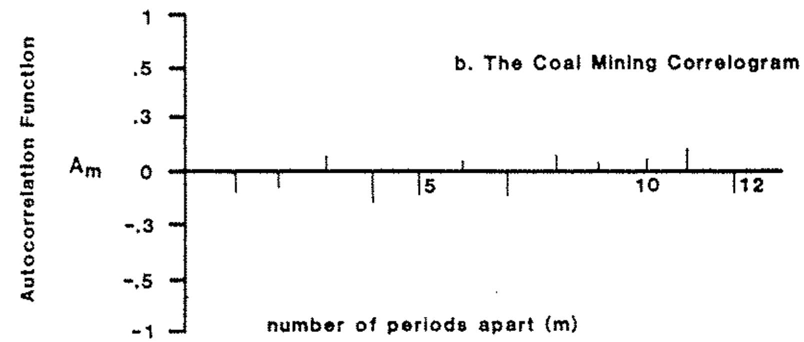


Figure 9.8 a, b: Empirical Autocorrelation Functions from the Residuals of the Interrupted Time Series Equations for Cuban Energy Consumption and United States Coal Mine Fatalities

autocorrelations, from $\hat{p}_1$, which correlates $\hat{e}_t$ with $\hat{e}_{t-1}$, to $\hat{p}_{12}$, which correlates $\hat{e}_t$ with $\hat{e}_{t-12}$. For the Cuban example, it means an empirical correlogram of 8 autocorrelation estimates, from $\hat{p}_1$ to $\hat{p}_8$. (To carry out these calculations, a special Box-Jenkins or ARIMA computer program is helpful, but not essential.) In Figures 9.8a–9.8b these empirical correlograms are shown.

In order to diagnose the particular error process, these empirical autocorrelations can be compared to theoretical ones, always bearing in mind that perfect correspondence cannot be expected. For example, by chance alone, one or two autocorrelation estimates might have some size, even though there was no autocorrelation in the population. Let us look first at the coal mine fatalities correlogram, which is more extended. First, the hypothesis of no significant first-order autocorrelation, earlier supported by the D-W statistic, is again confirmed. Clearly, no positive AR(1) process is at work, for there is exponential damping off of coefficients from $\hat{\rho}_1$. (Also, no positive second-order autoregressive process, which would exhibit steady, but not exponential, decay, is present.) With regard to negative autoregression, its systematic pattern of oscillation and decline is totally absent. Further, no MA(1) process is in evidence, as would be indicated by a spike at $\hat{\rho}_1$, followed by near zero coefficients. [An MA(2) process, which is also absent, would spike at $\hat{\rho}_2$ as well]. In sum, the autocorrelation function for the coal mine fatalities equation manifests no error dependency. Instead, it shows a random pattern. Note how all the coefficients are small, and scattered haphazardly around the zero line. Indeed, all 12 of the coefficients are far from statistical significance at .05. The implications of this "white noise" process are heartening. In particular, the OLS assumption of no autocorrelation is met. Therefore, the original OLS parameter estimates of equation 9.14 appear efficient, and the t-ratios valid. The causal inferences made from them have not been overturned by any discovery of an autocorrelation problem.

Turning to the Cuban case, the picture is somewhat less clear, in part because the empirical correlogram is shorter, going only to lag 8. This autocorrelation function reveals no obvious pattern of error dependency. Five negative and three positive autocorrelation estimates are distributed aimlessly around the zero horizon. Especially noteworthy is the fact that these coefficients are generally small. In fact, even the largest, $\hat{\rho}_7 = -.30$, is quite far from statistical significance at .05, given an effective sample size of 25 at that point. Again, the overall picture does not signal a violation of the no autocorrelation assumption. Hence the original OLS parameter estimates and t-ratios of the SITS model for Cuban energy consumption stand.

SUMMARY AND CONCLUSIONS

When the researcher wants to assess the effect of a relatively discrete event on a phenomenon observed across time, interrupted time-series analysis is applicable. Informally, the scatterplot is eyeballed to detect patterns of intervention. Formally, dummy variable time counters capturing the intervention are constructed and included in an ITS model. Then, the estimates record changes in the level (intercept) and trend (slope) of the time series that are associated with the intervention. The preferred estimation technique is OLS, provided the usual regression assumptions are met. ITS analysts must pay particular attention to the assumptions

of no specification error, no measurement error, no perfect multicollinearity, and no autocorrelation. Fortunately, however, violations of these assumptions are generally remediable, after which OLS can be applied. Briefly, here are some of the remedies already considered. Nonlinearity can usually be eliminated through variable transformation. Spuriousness can be controlled for by including other relevant variables. Improper measurement of the intervention point can be avoided by theoretical and empirical consideration of various lags. Perfect multicollinearity can be overcome by locating a proxy measure for the offending variable. First-order autocorrelation can be removed by generalized differencing, after which OLS is appropriate.Furthermore, by correcting one violation, another may also be corrected. For instance, accurate measurement of the intervention point avoids the specification error of including an irrelevant variable. Or, a remedy for nonlinearity may cure an autocorrelation problem. (Because the ITS variables are ordered by time, the D-W statistic can indicate nonlinearity, as well as autocorrelation; see Kmenta, 1971: 471.) Certainly, it is not possible to salvage the regression assumptions in every instance. Nevertheless, in general, OLS estimators for ITS models, at least after appropriate adjustments or transformation, clearly yield sound causal inferences. Of course, this prescription for using OLS, or a generalized version of it, may err. For instance, rarely, the problem may involve a mixed autoregressive-moving average error process, in which case a nonlinear estimation procedure may be preferred. However, such techniques are beyond the scope of this brief chapter.

references

Albritton, Robert B. (1979) "Measuring public policy: impacts of the Supplemental Social Security program." American Journal of Political Science 23: 559-578.

Ames, E. and S. Reiter (1961) "Distributions of correlation coefficients in economic time series." Journal of the American Statistical Association 56: 637-656.

Berry, William D. (1984) Nonrecursive Causal Models. Beverly Hills, CA: Sage.

———and Stanley Feldman (1985) Problems in Multiple Regression. Beverly Hills, CA: Sage.

Box, George E.P. and Gwilym M. Jenkins (1976) Time Series Analysis. San Francisco: Holden-Day.

Box, George E. P. and G. C. Tiao (1975) "Intervention analysis with applications to economic and environmental problems." Journal of the American Statistical Association 70: 70-92.

Campbell, D. T. (1969) "Reforms as experiments." American Psychologist 24: 409-429.

———and H. L. Ross (1968) "The Connecticut crackdown on speeding: time series data in quasi-experimental analysis. "Law and Society Review 3: 33-53.

Campbell, D. T. and J. C. Stanley (1966) Experimental and Quasi-Experimental Designs for Research. Skokie, IL: Rand McNally.

Caporaso, J. A. and A. L. Pelowski (1971) "Economic and political integration in Europe: a time series quasi-experimental analysis." American Political Science Review 65: 418-433.
Cook, T. D. and D. T. Campbell (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.
Deutsch, S. J. and F. B. Alt (1977) "The effect of Massachusetts' gun control law on gun-related crimes in the city of Boston." Evaluation Quarterly 1: 543-568.
Draper, N. R. and H. Smith (1966) Applied Regression Analysis. New York: John Wiley.
Eulau, Heinz and Michael S. Lewis-Beck (1985) Economic Conditions and Electoral Outcomes: The United States and Western Europe. New York: Agathon.
Glass, G. V (1968) "Analysis of data on the Connecticut speeding crackdown as a time series quasi-experiment." Law and Society Review 3: 55-76.
Hibbs, D. A., Jr. (1974) "Problems of statistical estimations and causal inference in time series regression models," in H. L. Costner (ed.) Sociological Methodology 1973-74. San Francisco: Jossey-Bass.
——— (1977) "On analyzing the effects of policy interventions: Box-Jenkins vs. structural equation models," in D. R. Heise (ed.) Sociological Methodology 1977. San Francisco: Jossey-Bass.
Kelejian, H. H. and W. E. Oates (1974) Introduction to Econometrics: Principles and Applications. New York: Harper & Row.
Kmenta, J. (1971) Elements of Econometrics. New York: Macmillan.
Lewis-Beck, M. S. (1979) "Some economic effects of revolution: models, measurement, and the Cuban evidence." American Journal of Sociology 84: 1127-1149.
——— (1980) Applied Regression: An Introduction. Beverly Hills, CA: Sage.
———(1981) "Can we assess the effects of revolution: a third look at the Cuban evidence." American Journal of Sociology 86: 1130-1133.
———and J. R. Alford (1980) "Can government regulate safety? The coal mine example." American Political Science Review 74: 745-756.
McCrone, Donald. J. and Richard J. Hardy (1978) "Civil rights policies and the achievement of racial income equality." American Journal of Political Science 22: 1-17.
McDowall, D., R. McCleary, E. E. Meidinger, and R. A. Hay, Jr. (1980) Interrupted Time Series Analysis. Beverly Hills, CA: Sage.
Nelson, J. P. and G. R. Neumann (1982) "Safety regulation and firm size: effects of the coal mine Health and Safety Act of 1969." Journal of Law and Economics 25: 183-199.
Ostrom, Charles W., Jr. (1978) Time Series Analysis: Regression Techniques. Beverly Hills, CA: Sage.
Perry, C. (1982) "Government regulation of coal mine safety: Effects of spending under strong and weak law." American Politics Quarterly 10: 303-314.
Pindyck, R. S. and D. L. Rubinfeld (1976) Econometric Models and Economic Forecasts. New York: McGraw-Hill.
Rao, P. and Griliches, Z. (1969) "Small-sample properties of several two-stage regression methods in the context of auto-correlated errors." Journal of the American Statistical Association 64: 253-272.
Tufte, E. R. (1974) Data Analysis for Politics and Policy. Englewood Cliffs, NJ: Prentice-Hall.
Zimring, F. E. (1975) "Firearms and federal law: The Gun Control Act of 1968." Journal of Legal Studies 4: 133-198.

10

Transfer Function Analysis

HELMUT NORPOTH

Like the federal government, the science of government lives by borrowing. Whereas the government borrows money to pay for many of its bills, political science borrows models to conduct much of its research. And, like the budget deficit in the public sector, resulting from a spendthrift government, there is something of an understanding deficit in the political science community. To many students of politics, the names of the borrowed concepts are unfamiliar and their meanings mystifying. The notations of the borrowed concepts deviate from familiar patterns; there are Greek characters to become familiar with in unfamiliar assignments. All that proves rather annoying if it turns out at the end that the new method is just old wine with a new label.

Transfer Function analysis is a recent addition to the methodological borrowing done by political scientists. The attempt to adopt this concept for the analysis of political data reveals all the troubles engendered by such borrowing. As introduced by Box and Jenkins (1976), transfer function analysis is largely geared to students of engineering and operations research.[1]

> Suppose X measures the level of an *input* to a system . . . the concentration of some constituent in the feed of a chemical process. Suppose that the level of X influences the level of a system *output* Y . . . the yield of product from the chemical process. It will usually be the case that, because of the inertia of the system, a change in X from one level to another will have no immediate effect on the output but, instead, will produce a delayed response with Y eventually coming to equilibrium at a new level. . . . A model which describes this dynamic response is called a *transfer function model* [Box and Jenkins, 1976: 335].

It is easy to think of inputs and outputs closer to home for most political scientists. One could take the rate of unemployment prevailing in the country as input to a system of voter evaluation of government, and the output of that system would be government popularity. Assume that unemployment has risen this month: By how much and over how many