DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Posc/Uapp 816

TIME SERIES ANALYSIS

I. AGENDA:
   A. Correction
   B. Time series
   C. Reading: Agresti and Finlay ***Statistical Methods in the Social Sciences***, 3rd edition, pages 537 to 538.

II. ANALYSIS OF THE COLEMAN REPORT:
   A. The notes for the last class (18) were in error regarding the analysis of the Coleman data.
      1. Here are the data properly labeled.

Staff	White collar	SES	Teacher Verbal	Mother's Education	Verbal
3.83	28.87	7.20	26.60	6.19	37.01
2.89	20.10	-11.71	24.40	5.17	26.51
2.86	69.05	12.32	25.70	7.04	36.51
2.92	65.40	14.38	25.70	7.10	40.70
3.06	29.59	6.31	25.40	6.15	37.10
2.07	44.82	6.16	21.60	6.41	33.90
2.52	77.37	12.70	24.90	6.86	41.80
2.45	24.67	-0.17	25.01	5.78	33.40
3.13	65.01	9.85	26.60	6.51	41.01
2.44	9.99	-0.05	28.01	5.57	37.20
2.09	12.20	-12.86	23.51	5.62	23.30
2.52	22.55	0.92	23.60	5.34	35.20
2.22	14.30	4.77	24.51	5.80	34.90
2.67	31.79	-0.96	25.80	6.19	33.10
2.71	11.60	-16.04	25.20	5.62	22.70
3.14	68.47	10.62	25.01	6.94	39.70
3.54	42.64	2.66	25.01	6.33	31.80
2.52	16.70	-10.99	24.80	6.01	31.70
2.68	86.27	15.03	25.51	7.51	43.10
2.37	76.73	12.77	24.51	6.96	41.01

      2. The estimated equation containing all of the variables is:

Verbal = 20.0 - 1.79 staff + 0.0439 Whtcollar + 0.555 SES + 1.11 Teachvrb - 1.82 Momed

3. It is **this** equation that contains some perhaps nonsensical results.
   i. The coefficient for "staff" is negative, when it should probably be positive, even when the other factors are taken into account.
      1) It's simple correlation with verbal scores is .192.
   ii. Similarly, the coefficient for mothers' education should probably be postive.
      1) The simple correlation is .733, a positive not negative relationship.
   iii. Moreover the partial verbal-mothers' education coefficient is not significant, although the two-variable correlation is highly significant.
4. Here are the more detailed regression results for the full model.

Predictor	Coef	StDev	T	P	VIF
Constant	20.01	13.64	1.47	0.164	
staff	-1.794	1.234	-1.45	0.168	1.4
Whtcolla	0.04386	0.05327	0.82	0.424	8.4
SES	0.55545	0.09296	5.97	0.000	3.5
Teachvrb	1.1100	0.4340	2.56	0.023	1.4
Momed	-1.822	2.029	-0.90	0.384	7.8

S = 2.075 R-Sq = 90.6%

   i. Only teachers' verbal scores and SES are significant.

B. We'll perhaps discuss these data further in class.

III. POLICY ISSUE: THE CAUSES OF WELFARE ONCE AGAIN:

A. Consider the previously discussed proposition advanced by Charles Murray. Murray suggests that the level of poverty in the United States declined until the 1970s when it began to level off and actually increase. He attributes this phenomenon in part to the growth of the welfare state (spending on anti-poverty programs). For example, he suggests that poverty was declining before President Johnson's Great Society and War on Poverty efforts were underway. Then, even as spending on social welfare increased, poverty slowly rose. Murray concludes that the welfare state has not arrested the increase in poverty, despite the billions of dollars spent in the attempt, and, if fact, may have actually worsened the situation.

B. In a recent article[1] he presents a variation on this theme: "But it is...important to confront the plain message of these data. At the same time that powerful social and economic forces were pushing down the incidence of black children born to

[1] Charles Murray, "Does Welfare Bring More Babies? The Public Interest Spring, 1994: 17-30.

unmarried couples, the incidence of black children born to unmarried women increased, eventually surpassing the rate for married couples. Something was making that particular behavior swim against a very strong tide, and, to say the least, <u>the growth of welfare is a suspect with the means and opportunity</u>."[2]

C. Once again "welfare"--be careful, Murray means by welfare the host of programs designed to aid the poor, not those aimed at the middle or upper classes, which incidentally consume a large part of the federal budget--contributes to or exacerbates an "undesirable" situation.

D. One might conceptualize this his as an example of an "intervention" that affected social and economic conditions. One could liken the enactment of poverty programs to a "quasi-experiment." Before the programs took effect (i.e., before the experiment got started) poverty was declining and illegitimacy was rising at a moderate rate. Then an intervention or interruption occurred. The question is: did the "intervention" (i.e., the enactment of the programs Murray decries) have an effect?
   1. Admittedly, this is not how he conceptualizes the problem. But on the other hand, it is entirely consistent with his (and related) arguments.

E. If Murray is correct, one would expect the to be "yes!," assuming other things were equal. Thus, what we can do is compare the pre- and post-intervention trends of appropriate dependent variables.
   1. See the attached see which gives the number of people living in poverty from 1959 to 1980.

IV. POLITICAL REGIMES AS INTERVENTIONS: THE CASE OF THE REAGAN ADMINISTRATION:

A. One may find it interesting to see what effect a change of political regime has on some factor that varies over time.
   1. Revolutions commonly produce such changes but even a stable system a change in the governing political party can potentially have significant effects as in the case of President Roosevelt's "New Deal" administration in the 1930s.
   2. In this vein it might be interesting to see what sort of effects President Reagan's administration produced.

B. Again, in order to keep matters simple, let's examine a couple of Reagan's most widely ballyhooed policy initiatives: the war on drugs and his <u>alleged</u> attack on organized labor. Because of the lack of data, we can't exhaustively explore this issue, but the data described later throw some light on how his administration affected American life.
   1. In particular, we will examine trends in
      i. Drug prosecutions and incarcerations
      ii. Strikes and union participation.

---

[2] Murray, p. 29. (Emphasis added.)

V. TIME SERIES:

A. Measurement of a variable at more or less equally spaced time intervals produces time series data, Y_t. The unit of analysis, in other words, is time. Measurements are taken at several time periods yielding a series of scores.

1. Examples: the rates of poverty, out-of-wedlock births, drug arrests, and strikes, to name only a few, increase or decrease over time.

B. The objective is to explain trends and changes in the series. We can use several methods:

1. Attempt to describe the nature of the change: does is represent a deterministic trend or random fluctuations.
2. Are there "seasonal" affects in addition to the trend and random variation that we observe?
3. Does the variation in the series change over time?
4. How are the observations at time t affected by previous values of t, previous values of an error, and/or previous values of one or more independent variables
5. Are there changes in the trend of a time series that can be interpreted or explained by an event or events?

C. **Interrupted Time Series (ITS)** analysis: since time-series analysis is a major statistical subject in its own right, we will deal with only one topic during next few classes, namely, how an **intervention** affects the time series.

D. The figure below represents the idea:

$$...Y_{t-3}, Y_{t-2}, Y_{t-1}, Y_t, \boldsymbol{I}\ Y_{t+1}, Y_{t+2}, Y_{t+3}...$$

Y's are measurements on the dependent at different times.
I is the "experimental" variable; that is, the event or "intervention" that supposedly "caused" a change in the time series.

Figure 1: Interrupted Time Series

E. The basic model that we will use is based on:

$$\boxed{Y_t = \beta_0 + \beta_1 t + \varepsilon_t}$$

where **t** is a dummy variable that indicates time periods (1,2,3...N, for N time periods)

i. Note also that the subscript, **t**, indicates the **tth** time period.

F. **Autoregression and serial correlated errors**: A major difficulty encounter in time series analysis is that the error term, ε_t, is frequently "serially" correlated. that is,

$$\boxed{\varepsilon_t = \rho\varepsilon_{t-1} + U_t}$$

1. This equation means that the error term at time **t** is a function of the error at the preceding time (**t - 1**), which as we saw earlier violates an OLS assumption. The effect of this "violation" is that the standard deviation of the regression coefficient will be too small and thus the t ratio too large, meaning that we will reject null hypotheses about **β**'s more frequently than we should.
2. Hence, we will see later how to measure and correct this assumption violation.

VI. VISUAL INSPECTION:

A. As always, the first step in analyzing time data is to try to draw pictures that both confirm preexisting and suggest new hypotheses.

B. Here are some figures for the data sets we will examine.

1. For this class, consider data relevant to the argument regarding regime change. The next figure shows the number of "regular" sentences of 60 months or more for drug violations meted out in federal district courts in the period 1945 to 1991.

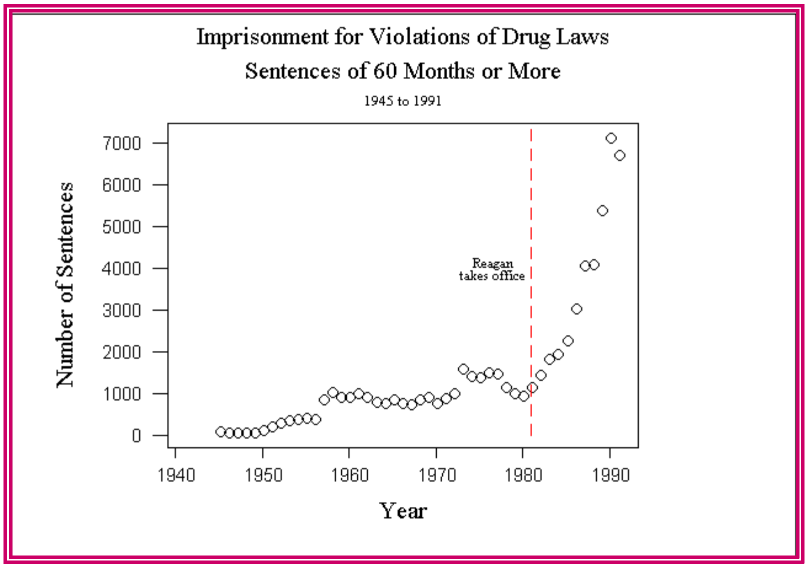

Figure 2: Time Series of Drug Sentences

2. The data suggest the following: the number of "severe" sentences grew

more or less gradually until President Reagan took office in 1981 when the number started to climb dramatically.

3. We face some interpretative difficulties:
   i. The growth could represent a dramatic increase in drug abuse.
   ii. It could represent a change in attitudes. (Recall Nancy Reagan's campaign against illegal drug use and the "epidemic" of cocaine addiction.) An alternative hypothesis, then, is that abuse did not increase nearly as rapidly as attention to the problem and efforts to deal with it.
   iii. Whatever the case, we an obvious first step is to compare the "slopes" of sentencing in the pre- and post-Reagan years.

C. Now let's look at labor union activity. The Reagan administration is credited with having broken the back of organized labor in the United States. On the other hand, union membership and strength had been declining for years before 1980. Thus, a natural question is did (and if so, to what degree) Reagan accelerate the process. The following data sheds some light on the question. It shows the trend in work "days lost" due to strikes and other labor activity.

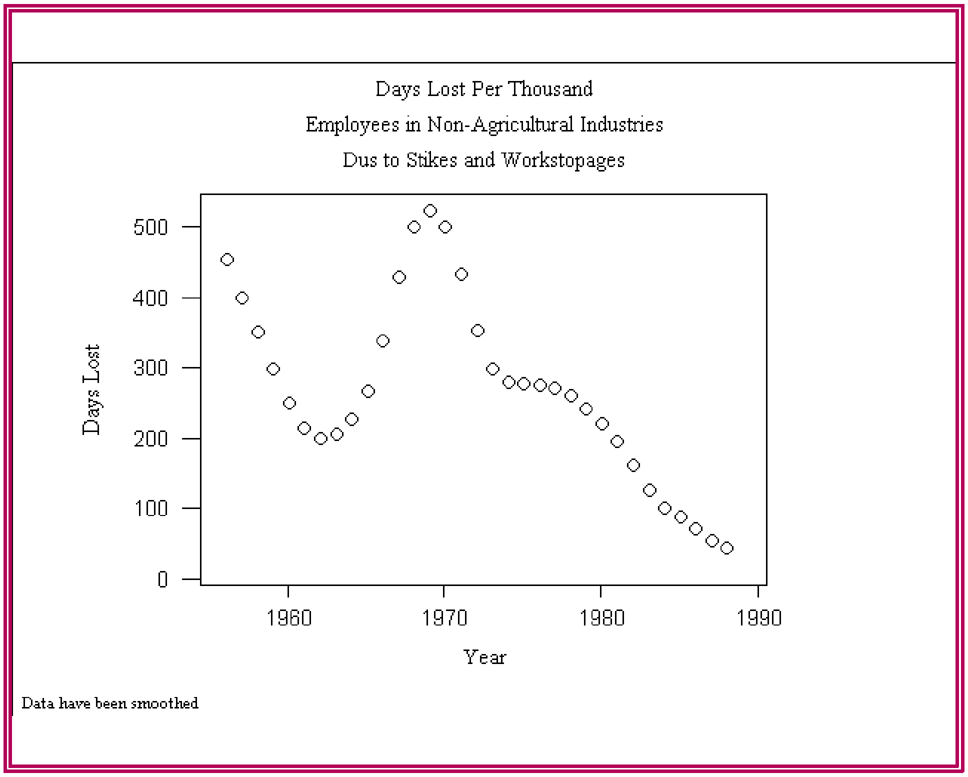

Figure 3: Work Stopages 1956 to 1988

VII. SIMPLE MODELS:

A. Here are some highly simple models that might describe the behavior of a time series.

B. Abrupt change in level, but no trend:

1. Consider this model where the level of the time series changes after an intervention.

$$Y_t = \beta_0 + \beta_1 X_1 + \varepsilon_t$$

***where*:**
X_1 = 0 for observations before I
X_1 = 1 for observations after I
I is the intervention
and is the error term

2. Here is a diagram of what such a model implies:

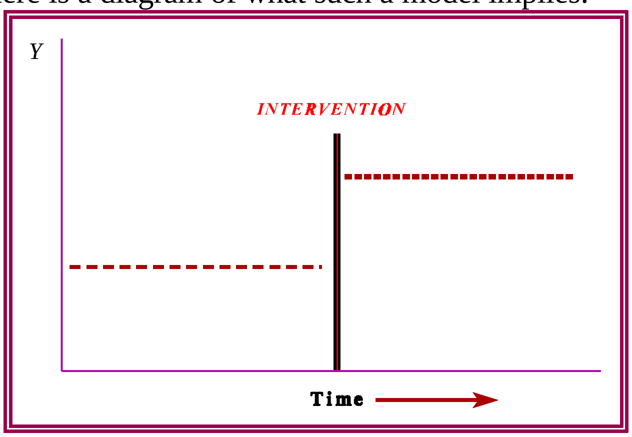

Figure 4: Change in Level

3. To see what this model implies, use our familiar technique of substituting in values of X_1 to see what results. When you do so, it will be apparent that the first coefficient is the level of Y_t before the intervention while the second is the effect of the intervention on the level. (In the above example, β_1 would be positive, since the level increases.

4. Note also that the change is permanent. We might want to develop a model in which the effect of the intervention died out over time.

C. Change (Permanent) in Trend:

1. Suppose the time series is trending downward at a sharp rate and then after the intervention starts to level off. A model that might capture this behavior is:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_t$$

where X_1 is a counter for time (t = 1,2...N)
X_2 = 0 for observations before I
$X_2 = X_1$ for observations on and after I
and ε_t is the error term

2. That is, X_2 is a counter variable for observations occurring after the intervention. Before, it equals 0.
3. A diagram of this model would be

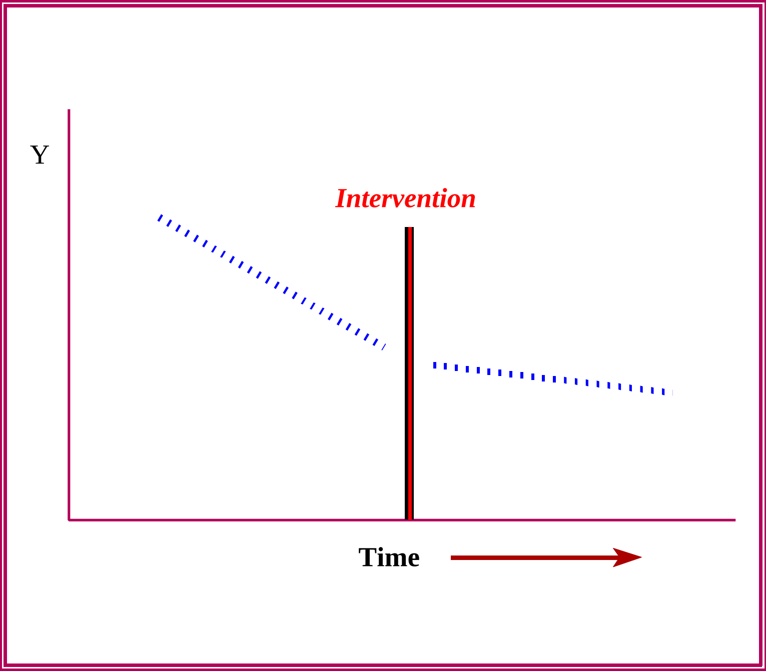

Figure 5: Change in Trend

4. To understand the equation, once again use the familiar technique of substituting values of X and time into the formula.
    i. This model suggest that the slope changes. In this particular case, the slope becomes less steep (less negative).
        1) Notice that the model posits a permanent change.
    ii. This model suggest that an intervention will create changes in both the level and slope of the time series.
    iii. Once again, the model can be made more understandable by making the appropriate substitutions. For example:
        1) Pre-intervention:

$$E(Y_t) = \beta_0 + \beta_1 X_1$$
$$\textit{because } X_2 = 0$$

2) The parameter β_1 is the slope of the trend in the pre-intervention period.

iv. Post-intervention:

$$E(Y_t) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$
$$= \beta_0 + (\beta_1 + \beta_2) X_1$$
$$\textit{because } X_2 = X_1 \textit{ on/and after post-intervention}$$

v. Look at the definitions of the variables to convince yourself of these relationships.

vi. β_2 is the adjustment to the trend or slope after the intervention.

D. Change in level and slope.

1. To model a change in both the slope and level of a trend we can add a dummy variable coded 0 for observations before the intervention and 1 after:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_t$$

where X_1 *is a counter for time* $(t = 1,2...N)$
$X_2 = 0$ *for observations before I and 1 after*
$X_3 = 0$ *for observations before I*
$X_3 = X_1$ *for observations on and after I*
and ε_t *is the error term*

i. Once again substitute to see what the parameters mean.

1) Pre-intervention $X_2 = X_3 = 0$ and

$$E(Y_t) = \beta_0 + \beta_1 X_1$$

2) After the intervention, X_2 = 1 and X_3 = X_1, so

$$E(Y_t) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$
$$= \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1$$
$$= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$$
$$= \beta_0^* + \beta_1^* X_1$$

where β_0^* and β_1^* are the adjusted parameters

2. The model asserts that there is a change in the level and trend of the time series.
   i. The level or constant is adjusted by β_2 and the trend or slope by β_3.

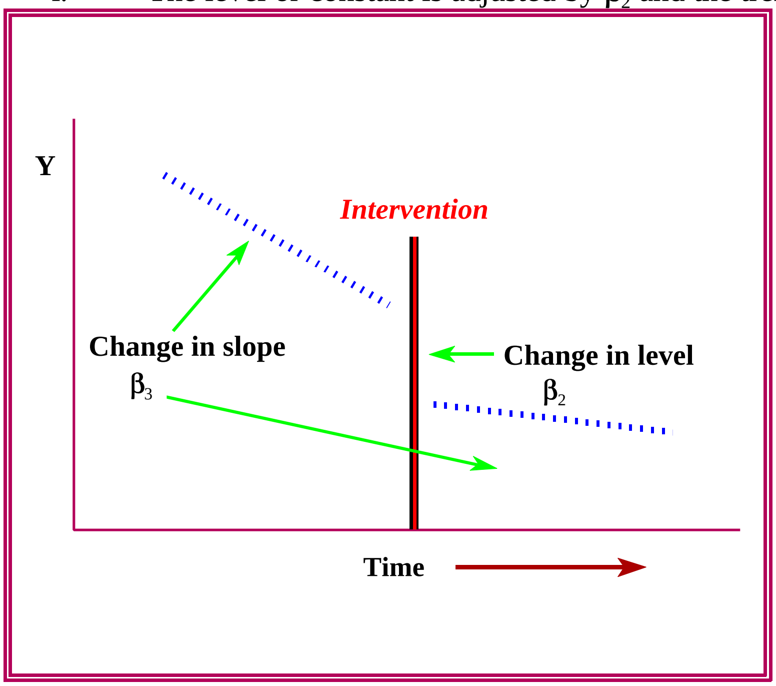

Figure 5: Change in Trend

   ii.

E. We'll explore these models in more detail next time.

VIII. NEXT TIME:
A. Time series and intervention analysis.

Go to Notes page

Go to Statistics page