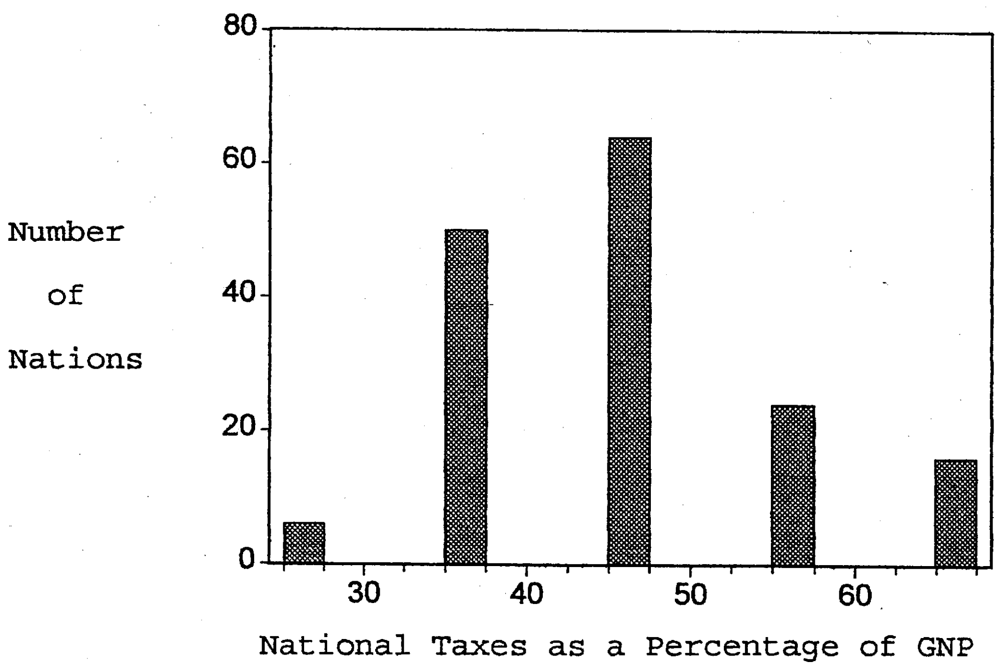
The table below is the frequency distribution of international tax rates from page 15. The graph on the lower portion of this page is a histogram of the same data as appear in the frequency distribution.

Frequency Distribution of International Tax Rates
(Hypothetical Data)

National Taxes as a Percentage of National GNP	Percentage of Nations	Number of Nations
70% or more	0%	0
60 - 69	10	16
50 - 59	15	24
40 - 49	40	64
30 - 39	31	50
20 - 29	4	6
0 - 19	0	0
	100%	160

Histogram of International Tax Rates
with Data at the Midpoint of the Interval
(i.e., all 16 nations having national tax rates between 60 and 69 percent of GNP are list as 65 percent)



Interpreting the Mean and Standard Deviation

The table below contains the mean and standard deviation for three nations on a quality of life indicator. The quality of life indicator is a 10 point scale with "1" indicating the poorest quality of life and "10" indicating the highest quality of life.

Spain		Japan		United States	
Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
4.9	2.7	7.1	1.1	6.4	1.3

Suppose you were an official with the World Bank. Let us say that each of these three nations applied for a loan. If you could only loan money to one nation, which one would you select? Why? [redacted]

Variable Associations

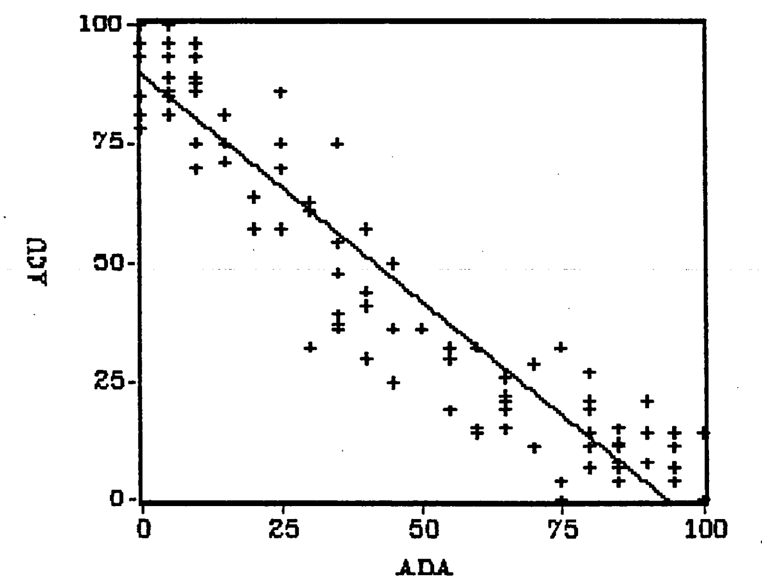
Observation	X	Y
Observation #1	1	7
Observation #2	3	5
Observation #3	5	3
Observation #4	7	1

Make sure you can distinguish the scores on variables from a measure which associates the scores on variables. The example above shows a perfect negative correlation between X and Y (i.e., -1.0). Are any of the above scores negative?

LS // Dependent Variable is ACU
Date: 4-08-1996 / Time: 4:06
SMPL range: 1 - 100
Number of observations: 100

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	89.826047	1.8026615	49.829680	0.0000
ADA	-0.9588531	0.0320857	-29.884143	0.0000

R-squared	0.901116	Mean of dependent var	46.39000
Adjusted R-squared	0.900107	S.D. of dependent var	33.73722
S.E. of regression	10.66293	Sum of squared resid	11142.41
Log likelihood	-377.5610	F-statistic	893.0620
Durbin-Watson stat	1.742154	Prob(F-statistic)	0.000000



Bivariate Regression and Scatter Plot of Conservatism (ACU) being explained by Liberalism (ADA)

Figure 1

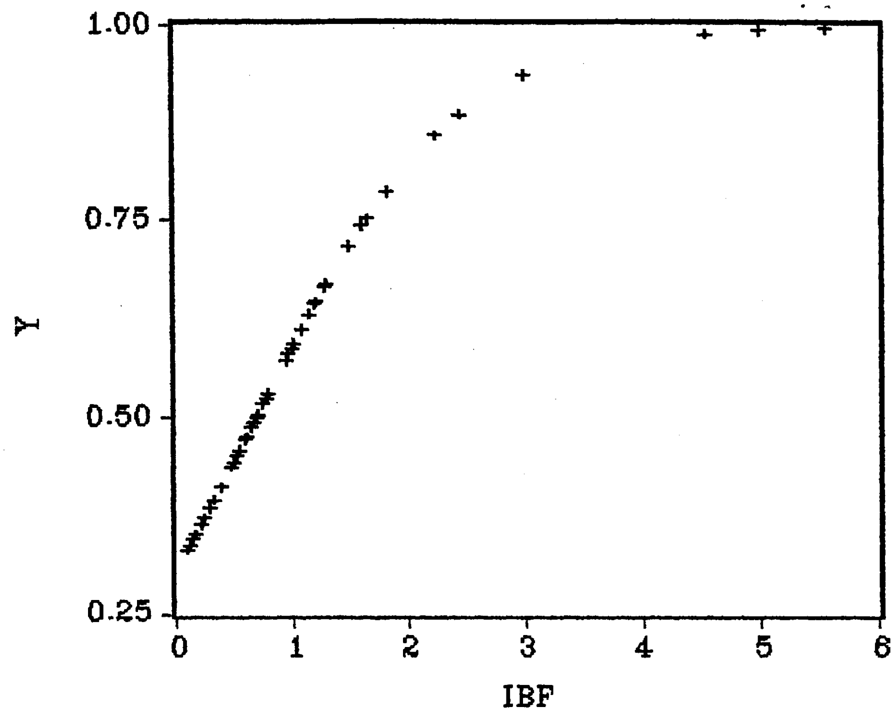


Figure 2

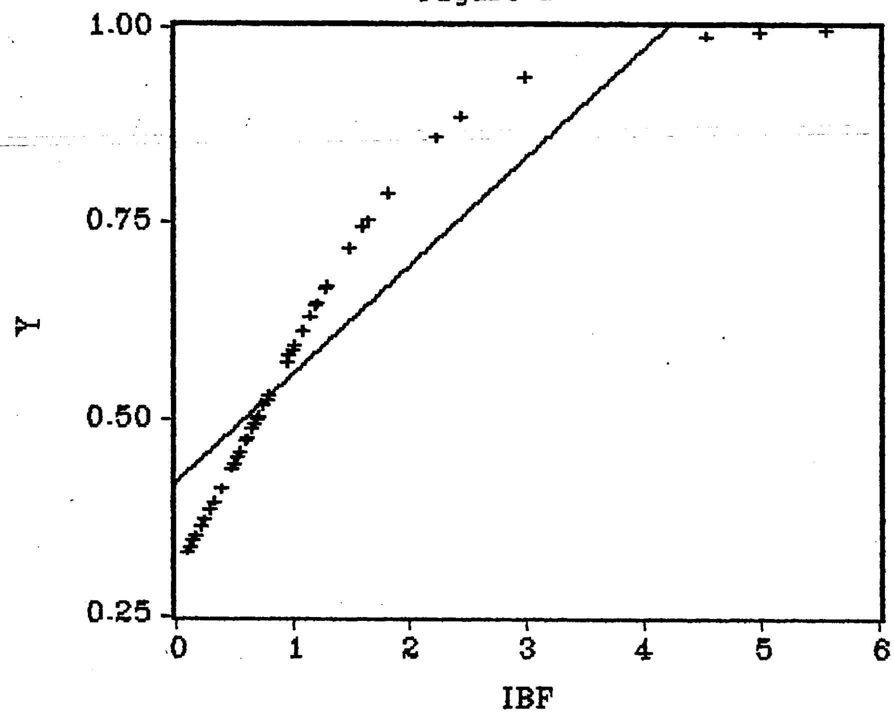


Figure 1 shows the probability (calculated from a logit model) that deterrence will be successful (Y) by the ratio of immediately deployable forces of the defender and the protege to those of the potential attacker (IBF) for the 58 observations in Assignment 3 in POSC 300B. Figure 2 is the same as Figure 1 except that a regression line has been added.

As discussed over pages 106-114, political scientists use multiple (i.e., more than one independent variable) regression much more frequently than bivariate (i.e., one independent variable) regression because few dependent variables are influenced only by one independent variable and because the impact of each independent variable on a particular dependent variable is likely to change depending upon which independent variables are in the equation. The following table shows the results for the impact of a senator's conservatism (CONS), political party affiliation (PARTY) and state median family income (STINC) on the senator's support for the poor on tax legislation (TAX) estimated by both multiple regression (i.e., the effect of all three independent variable estimated in one equation: $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e$) and by three bivariate equations (each with one independent variable: $Y = a + b_1X_1 + e$; $Y = a + b_2X_2 + e$ and $Y = a + b_3X_3 + e$). This example is from Assignment #2 in POSC 300B.

	Multiple Regression		Bivariate Regression	
	Coefficient (b)	T Statistic	Coefficient (b)	T Statistic
CONS	-.644	-8.52	-.737	-13.279
PARTY	11.207	2.39	35.293	7.409
STINC	-.560	-.43	2.867	1.523

Notice the difference in the results. While the coefficient (i.e., b) for the senator's conservatism (CONS) is somewhat different (-.644 vs. -.737) the t statistics are quite different (-8.52 vs -13.279). The differences are much more apparent for the senator's party affiliation (PARTY). In the multiple regression results, the impact of party is less than one-third of the impact in the bivariate equation (11.207 vs. 35.293). This occurs because conservatism and political party affiliation are highly related. Typically, Republican senators are much more conservative than Democratic senators (i.e., multicollinearity - your next reading assignment). So, when the effect of party is estimated without taking into account conservatism, party receives much of the impact of conservatism. Conservatism is not effected that much by the omission of party because the impact that conservatism has on the dependent variable that it does not also share with party is about the same as the impact that conservatism has if party is omitted (more on this when we discuss next week's reading assignment). Notice that both the absolute size of the coefficient and the sign (positive or negative) for state median family income in the senator's state (STINC) changes depending upon whether we are using multiple or bivariate regression. Assuming we have a logical model, multiple regression produces more accurate estimates of the impact of each of the independent variables than bivariate regression. This is why political scientists invariably build multivariate models (i.e., use multiple regression).

Pages 250-253 show you how the computer estimates the value of each "b" in a multiple regression equation. Thus, you will see how equations 4-6 over pages 113 and 114 are executed. We begin with equation 4 on page 113. The purpose of this equation is to obtain that portion of the senator's conservatism (X_1) that cannot be explained by either the senator's political party affiliation (X_2) or the median income in the senator's state (X_3). The portion of the senator's conservatism that cannot be explained by either party affiliation or state median family income is E2 from equation 4 on page 113.

LS // Dependent Variable is CONS→X_1
Date: 10/10/98 Time: 18:52
Sample: 1 100
Included observations: 100

Equation 4
Page 113

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C→a_2	137.4466→a_2	14.74029	9.324551	0.0000
PARTY→X_2	-41.13872→b_4	4.688349	-8.774670	0.0000
STINC→X_3	-8.346615→b_5	1.500567	-5.562307	0.0000

R-squared	0.493782	Mean dependent var	35.11000
Adjusted R-squared	0.483345	S.D. dependent var	31.24258
S.E. of regression	22.45677	Akaike info criterion	6.252725
Sum squared resid	48917.73	Schwarz criterion	6.330880
Log likelihood	-451.5301	F-statistic	47.30859
Durbin-Watson stat	1.614248	Prob(F-statistic)	0.000000

Below are the values of e_2 from equation 4 on Page 113

E2

Last updated: 10/10/98 - 19:29
Modified: 1 100 // e2=resid

1	-8.542927	-17.54293	8.457073	-20.54293	5.662259
6	-25.70827	-23.20428	-15.34571	-8.345714	4.795720
11	5.843531	-3.345714	21.19015	-15.34571	6.182182
16	-9.180371	-5.180371	6.000901	4.000901	34.00090
21	-16.36165	-10.36165	-6.023008	-18.02301	30.19015
26	-9.676392	-13.67639	-0.495112	-6.164434	-27.86564
31	-20.86564	-21.18834	-17.18834	-7.841728	-12.84173
36	-4.495112	-17.19632	8.512860	-33.70827	16.99294
41	7.976992	-5.007063	-1.503085	2.496915	-22.02301
46	-10.68436	0.126394	15.79572	8.275797	-25.72420
51	13.60647	21.60647	-9.377588	38.62241	14.29173
56	12.29173	37.13436	40.13436	4.969028	16.96903
61	71.81166	12.63835	-9.342998	0.523540	-1.342998
66	21.98768	20.65700	-41.97247	16.02753	-27.31909
71	-49.31909	-17.15375	-19.15375	-6.177663	22.82234
76	-35.29519	-6.956537	38.86218	22.33430	29.33430
81	39.34226	34.34226	4.051428	23.36617	-51.47646
86	-29.79917	-1.799166	7.318355	-21.68165	-34.83105
91	2.168952	-22.63383	21.66497	7.374142	23.05143
96	24.83827	21.67294	-44.16173	24.17692	14.49963

Just to make sure you understand how the values for e_2 on page 250 are calculated, let us review the procedure for predicting scores on the dependent variable (in this case, X_1, the senator's conservatism or "CONS") from scores on the independent variables (in this case, X_2, the senator's party affiliation or "PARTY" and X_3, the median family income in the senator's state or "STINC"). From equation 2 on page 111, we know that the formula to predict scores on X_1 from X_2 and X_3 is:

$$\hat{X}_1 = a_2 + b_4X_2 + b_5X_3$$

Verbally, the equation immediately above says to the computer, using the least squared errors method, make the most accurate prediction of a senator's conservatism you can based upon a knowledge of the senator's political party affiliation and the median family income in the senator's state. From the results on the top of page 250, we know that $a_2 = 137.447$, $b_4 = -41.139$ and $b_5 = -8.347$. From the top of page 197, we know that senator #1's scores on the variables of interest are: PARTY = 1 and STINC = 7.4. Substituting these values into the above formula yields:

$$\hat{X}_1 = 137.447 + [(-41.139)(1)] + [(-8.347)(7.4)$$

$$= 137.447 + (-41.139) + (-61.768)$$

$$= 137.447 - 41.139 - 61.768$$

$$= 34.54$$

From the top of page 197, we know that senator #1's actual score on conservatism is 26. From page 80, we know that the formula for the residual or in this case, "e_2", is:

$$e_2 = X_1 - \hat{X}_1$$

Since $\hat{X}_1$ represents the best prediction we can make of the level (amount) of conservatism of a senator based upon knowledge of the senator's political party affiliation and the median family income in the senator's state, the error term, e_2, must represent that portion of a senator's conservatism which cannot be predicted from a knowledge of the senator's political party affiliation and the median family income in the senator's state. Put another way, e_2 represents X_1 after the effects of party affiliation and median family income have been removed, or "controlled." Substituting our estimate of X_1 for senator #1, 34.54, and the actual score on X_1 for senator #1, 26, into the above formula yields:

$$e_2 = 26 - 34.54$$

$$= -8.54$$ (which is the first score for e_2 on page 250)

This result, -8.54, is that portion of senator #1's conservatism that cannot be explained by the senator's party affiliation and the state median family income in their state. In order to obtain the 100 scores for e_2, the computer would then repeat the process we just went through for the remaining 99 senators. In order to estimate b_1 in equation 3 on page 112 (the impact of the senator's conservatism on Y - the percentage of times the senator supported the poor on tax legislation - after removing the effects of X_2 and X_3), we will relate that portion of X_1 that cannot be explained by X_2 and X_3 (i.e., after X_2 and X_3 have been "controlled," hence e_2) to the portion of Y that cannot be explained by X_2 and X_3 (i.e., after X_2 and X_3 have been "controlled," hence e_3 on the next page). The scores for the portion of Y that cannot be explained by X_2 and X_3, e_3, are on page 252.

LS // Dependent Variable is TAX→Y
Date: 10/10/98 Time: 18:54
Sample: 1 100
Included observations: 100

Equation 5
Page 113

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C→a_3	-21.23187→a_3	14.47513	-1.466783	0.1457
PARTY→X_2	37.73090→b_6	4.604012	8.195221	0.0000
STINC→X_3	4.821153→b_7	1.473574	3.271741	0.0015

R-squared	0.422789	Mean dependent var	46.54000
Adjusted R-squared	0.410888	S.D. dependent var	28.73193
S.E. of regression	22.05280	Akaike info criterion	6.216420
Sum squared resid	47173.63	Schwarz criterion	6.294575
Log likelihood	-449.7149	F-statistic	35.52473
Durbin-Watson stat	1.733571	Prob(F-statistic)	0.000000

Below are the values of e_3 from equation 5 on page 113

E3

Last updated: 10/10/98 - 19:30
Modified: 1 100 // e3=resid

1	1.824440	10.82444	-11.17556	-27.17556	-1.156905
6	44.34233	9.895979	-0.264211	-10.26421	-21.10402
11	-19.46017	23.73579	8.718676	13.73579	-18.38864
16	28.21790	21.21790	-21.08536	-3.085365	-20.08536
21	26.52117	32.52117	-2.407287	22.59271	-52.28132
26	19.77156	12.77156	24.46829	23.43252	22.96752
31	34.96752	36.11060	37.11060	14.28944	7.289442
36	12.46829	8.003289	-47.42441	39.34233	26.80732
41	-35.40729	23.80732	-5.639018	-22.63902	23.59271
46	5.664248	19.86021	-24.10402	-22.87229	28.12771
51	-20.90806	-20.90806	-29.69344	-26.69344	-31.65767
56	-31.65767	-35.03248	-10.03248	-12.51460	0.485402
61	-43.88940	-36.47883	-0.373124	0.573992	-7.373124
66	-8.408892	-8.373124	23.12765	-0.872354	-14.05120
71	39.94880	-8.569084	9.430916	12.10899	0.108993
76	29.27073	-0.657738	-16.35447	-13.23005	-10.23005
81	-1.122737	-11.12274	-22.55043	-18.80081	32.57399
86	22.71707	-15.28293	11.55534	-3.444662	4.287840
91	3.287840	7.199186	-2.265813	-24.69351	2.449573
96	-9.676390	-7.158507	31.32361	-12.60485	-12.74793

The results below come from asking the computer to use the least squared errors procedure for estimating the following equation (equation 6 from the top of page 114):

$e_3 = a_4 + b_1 e_2 + e_4$ where e_2 and e_3 are as generated from the results on pages 250 and 252 and b_1 is the value of b_1 in equation 3 on page 112: $Y = a_1 + b_1 X_1 + b_2 X_2 + b_3 X_3 + e_1$

Notice that the value of b_1 below (-.644) is identical to the value of b_1 in equation 1 on page 195.

LS // Dependent Variable is E3
Date: 10/10/98 Time: 19:31
Sample: 1 100
Included observations: 100

Equation 6
Page 114

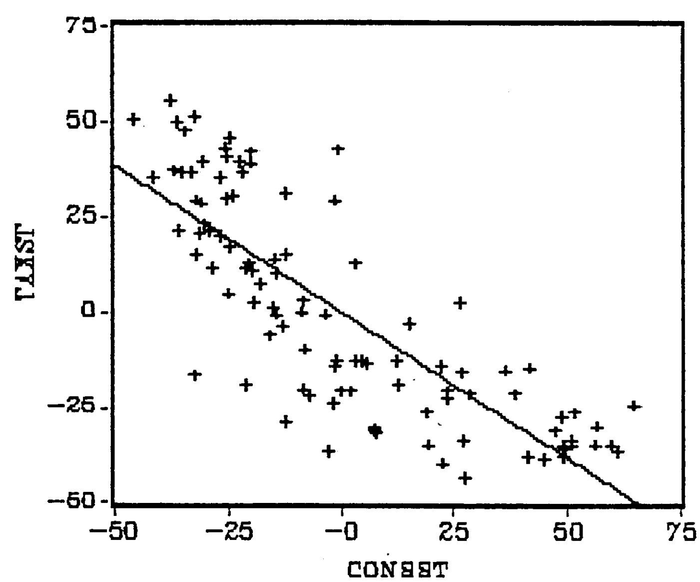
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C → a_4	5.73E-15	1.654933	3.46E-15	1.0000
E2 → b_1	-0.644721	0.074825	-8.616369	0.0000

R-squared	0.431032	Mean dependent var	-1.72E-15
Adjusted R-squared	0.425227	S.D. dependent var	21.82891
S.E. of regression	16.54933	Akaike info criterion	5.632488
Sum squared resid	26840.26	Schwarz criterion	5.684592
Log likelihood	-421.5183	F-statistic	74.24182
Durbin-Watson stat	2.019803	Prob(F-statistic)	0.000000

LS // Dependent Variable is TAXST
Date: 4-07-1996 / Time: 4:52
SMPL range: 1 - 100
Number of observations: 100

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	2.636E-08	1.7037469	1.547E-08	1.0000
CONSST	-0.7652795	0.0575161	-13.305478	0.0000

R-squared	0.643683	Mean of dependent var	1.59E-08
Adjusted R-squared	0.640047	S.D. of dependent var	28.39763
S.E. of regression	17.03747	Sum of squared resid	28446.98
Log likelihood	-424.4252	F-statistic	177.0357
Durbin-Watson stat	1.954313	Prob(F-statistic)	0.000000

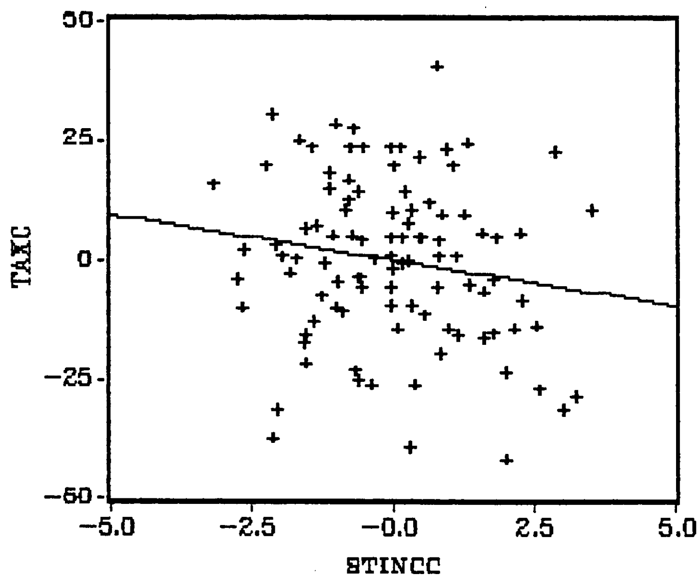


Comparing the graphs and t statistics on pages 254 and 255 show the following: the higher the absolute value of the t ratio the steeper the regression line, the closer the dots surround (or "fit") the line and the more variation (i.e., greater difference in scores) on the independent variable.

LS // Dependent Variable is TAXC
Date: 4-07-1996 / Time: 5:12
SMPL range: 1 - 100
Number of observations: 100

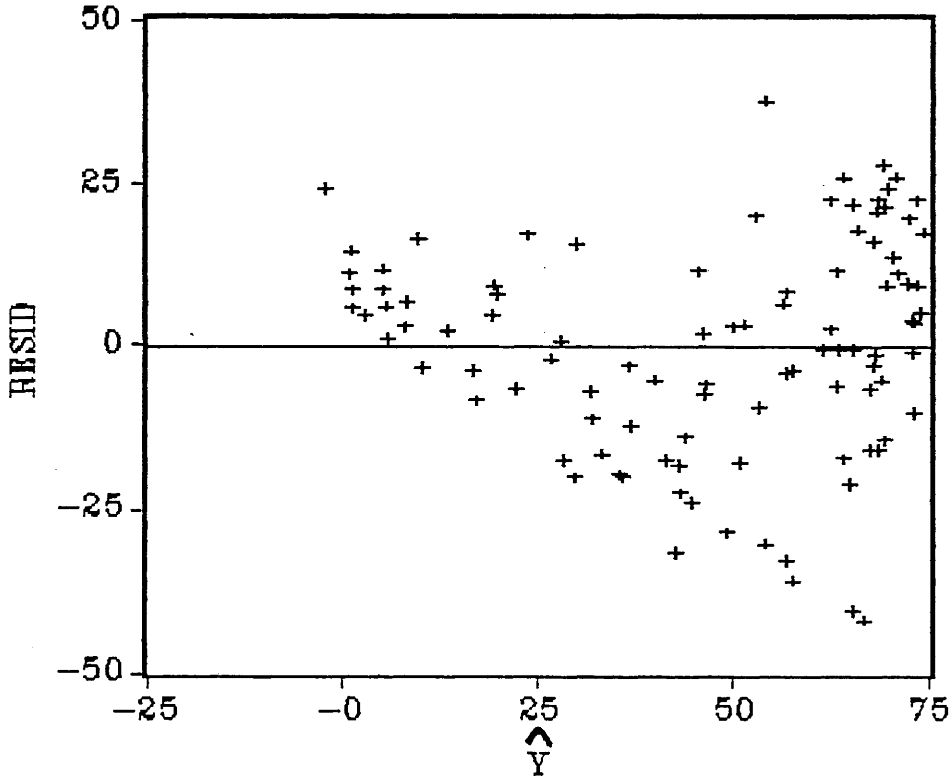
VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-2.405E-08	1.7037469	-1.412E-08	1.0000
STINCC	-1.8899102	1.1789680	-1.6030208	0.1121

R-squared	0.025551	Mean of dependent var	-2.89E-08
Adjusted R-squared	0.015608	S.D. of dependent var	17.17200
S.E. of regression	17.03747	Sum of squared resid	28446.98
Log likelihood	-424.4252	F-statistic	2.569676
Durbin-Watson stat	1.954313	Prob(F-statistic)	0.112148



Bivariate Regression and Scatterplot for "Tax" (with the effects of "Cons" removed) being explained by "Stinc" (with the effects of "Cons" removed) - from Assignment #2 in POSC 300B

256



The diagram above shows a plot of the residuals from Assignment #2 in POSC 300B ("RESID" - which is on the vertical axis) by the predicted values for Y (i.e., $\hat{Y}$ - which is on the horizontal axis). Do the residuals appear heteroscedastic?

White Heteroskedasticity Test:

Test Equation:
LS // Dependent Variable is RESID^2
Date: 06/04/97 Time: 11:29
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3952.827	2261.959	1.747524	0.0839
CONS	-13.49558	18.34115	-0.735809	0.4637
CONS^2	-0.079158	0.060584	-1.306572	0.1947
CONS*PARTY	-0.628181	4.324808	-0.145251	0.8848
CONS*STINC	2.230318	1.511155	1.475903	0.1434
PARTY	-1144.282	909.8235	-1.257696	0.2117
PARTY*STINC	143.9888	84.05957	1.712938	0.0901
STINC	-685.2522	383.9091	-1.784933	0.0776
STINC^2	28.67171	16.87565	1.698999	0.0927

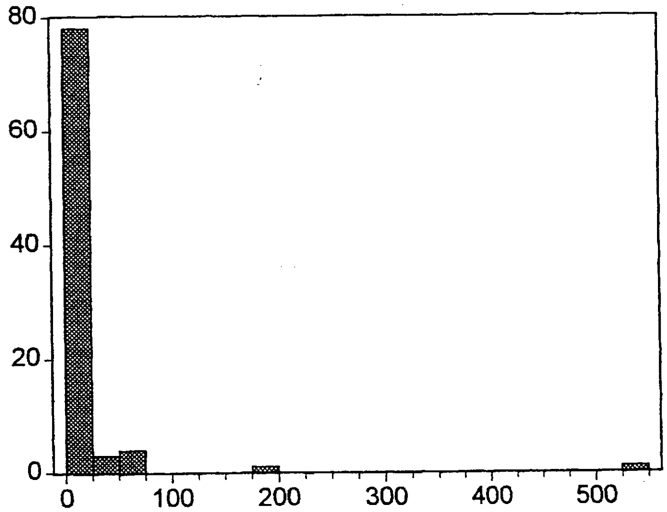
R-squared	0.170871	Mean dependent var	268.4026
Adjusted R-squared	0.097980	S.D. dependent var	355.3311
S.E. of regression	337.4747	Akaike info criterion	11.72867
Sum squared resid	10363913	Schwarz criterion	11.96314
Log likelihood	-719.3274	F-statistic	2.344213
Durbin-Watson stat	2.110998	Prob(F-statistic)	0.024435

LS // Dependent Variable is TAX
Date: 06/04/97 Time: 21:32
Sample: 1 100
Included observations: 100
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67.38277	17.22732	3.911390	0.0002
CONS	-0.644721	0.063158	-10.20802	0.0000
PARTY	11.20792	4.298106	2.607642	0.0106
STINC	-0.560081	1.530482	-0.365951	0.7152

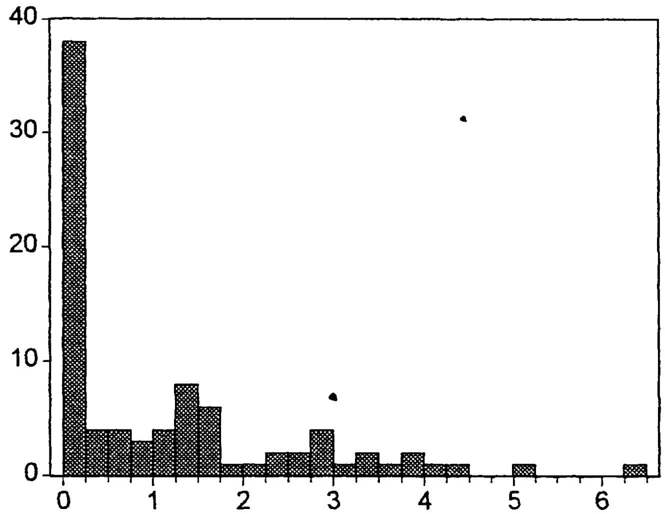
R-squared	0.671586	Mean dependent var	46.54000
Adjusted R-squared	0.661323	S.D. dependent var	28.73193
S.E. of regression	16.72083	Akaike info criterion	5.672488
Sum squared resid	26840.26	Schwarz criterion	5.776695
Log likelihood	-421.5183	F-statistic	65.43790
Durbin-Watson stat	2.019803	Prob(F-statistic)	0.000000

The first equation above shows the White test for heteroscedasticity applied to equation 1 (page 195) of Assignment #2 in POSC 300B. Notice that while some of the "t-statistics" are close to 2.0, none actually attain it. The second equation above shows that after adjusting for heteroscedasticity (see page 139) the standard errors of the coefficients are lower, and the "t-statistics" higher, than those reported for equation 1 on page 195.



Series: DPM75
Sample 1 87
Observations 87

Mean	15.34424
Median	1.737101
Maximum	546.2951
Minimum	1.000000
Std. Dev.	61.82816
Skewness	7.583422
Kurtosis	63.80187
Jarque-Bera	14235.02
Probability	0.000000



Series: LDPM75
Sample 1 87
Observations 87

Mean	1.145460
Median	0.552217
Maximum	6.303159
Minimum	0.000000
Std. Dev.	1.400793
Skewness	1.353088
Kurtosis	4.412065
Jarque-Bera	33.77526
Probability	0.000000

Histogram of Violent Political Acts over 1973-75 (DPM75) and the Natural Logarithm of Violent Political Acts over 1973-75 (LDPM75). Political Violence is measured by the death rate from domestic conflict (i.e., political-protest events such as riots, armed attacks and assassinations) per one million people. Expressing scores on a variable in logarithms is especially useful when the scores are badly skewed (i.e., not symmetrical like the normal curve on page 24). Notice how the scores on the extreme right side of the top graph are very isolated from the other scores. These extreme scores are called "outliers." Now notice that in the lower graph the scores are more dispersed (i.e., spread out) over the entire range between the lowest and highest scores (i.e., the extreme scores are not so "outlying" in the lower graph). Converting to logarithmic units reduced the skewness.

From page 24 we know that with a normal distribution, the mean, median and mode are all at the same point. Notice in the top graph how much difference there is between the mean and median (15.3 vs. 1.73). Now notice that in the lower graph, with a logged variable, the mean and median are much closer (1.1 vs. .5). This suggests that the scores in the lower graph are closer to being normally distributed than the scores in the upper graph. The Jarque-Bera test is a test to determine how likely a sample distribution of scores is to have come from a normal distribution. While the mathematics of the test are somewhat involved, the probabilities immediately below the Jarque-Bera test results tell how likely we would have the distribution of scores in our sample, if the population these scores were drawn from was normal. If the probability is .05, or lower, we reject the null hypothesis that the sample scores come from a normal distribution because we have a 5%, or less, chance that the scores actually come from a normal distribution. Obviously, while the scores in the lower graph are closer to a normal distribution than the scores in the upper graph, neither graph is close to a normal distribution (since both probabilities are .000000, easily below the .05 threshold).

LS // Dependent Variable is DPM75
Date: 10/08/98 Time: 17:56
Sample: 1 87
Included observations: 57
Excluded observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-211.6113	101.3842	-2.087221	0.0417
RUPPER20	2.689784	1.237775	2.173079	0.0343
LENERGY	12.14195	8.431451	1.440078	0.1557
RSEMI	33.46578	23.51587	1.423115	0.1606

R-squared	0.136775	Mean dependent var	19.33852
Adjusted R-squared	0.087914	S.D. dependent var	75.47829

LS // Dependent Variable is LDPM75
Date: 10/08/98 Time: 17:57
Sample: 1 87
Included observations: 57
Excluded observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.257922	1.830885	-1.233241	0.2229
RUPPER20	0.037974	0.022353	1.698859	0.0952
LENERGY	0.149858	0.152263	0.984208	0.3295
RSEMI	1.327722	0.424670	3.126478	0.0029

R-squared	0.261786	Mean dependent var	1.230425
Adjusted R-squared	0.220000	S.D. dependent var	1.473953

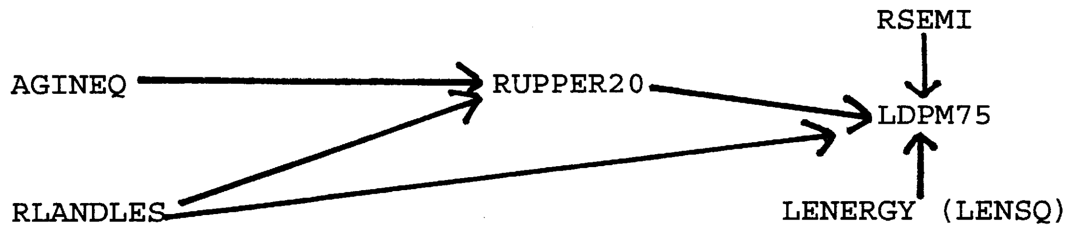
LS // Dependent Variable is LDPM75
Date: 10/08/98 Time: 17:58
Sample: 1 87
Included observations: 57
Excluded observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.142474	2.991097	-1.384935	0.1720
RUPPER20	0.031784	0.023732	1.339312	0.1863
LENERGY	0.901748	0.954050	0.945179	0.3489
LENSQ	-0.059607	0.074657	-0.798408	0.4283
RSEMI	1.250376	0.437003	2.861252	0.0061

R-squared	0.270726	Mean dependent var	1.230425
Adjusted R-squared	0.214628	S.D. dependent var	1.473953

Multiple Regression of Violent Political Acts over 1973-75 (DPM75) and the Natural Logarithm of Violent Political Acts over 1973-75 (LDPM75) explained by The Share of National Income going to the Richest 20% of the Population, the Natural Logarithm of Per Capita Energy Consumption, the Square of the Natural Logarithm of Per Capita Energy Consumption, whether the Regime is Semi-Repressive or not (1 = semi-repressive; 0 = either very repressive - e.g., Adolph Hitler or a non-repressive regime - e.g., United States)

The results on page 259 show us the direct impact of the percentage of income going to the richest 20% of the population (RUPPER20), the level of economic development (LENERGY and LENSQ) and whether or not the regime is semi-repressive (RSEMI) on the level of political violence (LDPM75). A richer explanation can be obtained from using the causal model drawn below which is estimated from the regression results immediately below the diagram. AGINEQ is the degree of agricultural inequality in the nation and RLANDLES is the percentage of the population in the nation that does not own land.



LS // Dependent Variable is RUPPER20
Date: 10/13/98 Time: 12:14
Sample: 1 87
Included observations: 45
Excluded observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.97820	2.377571	14.29114	0.0000
AGINEQ	3.252471	0.692783	4.694792	0.0000
RLANDLES	0.028982	0.107231	0.270275	0.7883

R-squared	0.557257	Mean dependent var	48.38444
Adjusted R-squared	0.536174	S.D. dependent var	8.099774

LS // Dependent Variable is LDPM75
Date: 10/13/98 Time: 12:30
Sample: 1 87
Included observations: 44
Excluded observations: 43

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.798922	4.542334	-1.496791	0.1427
RUPPER20	0.011374	0.030232	0.376210	0.7089
LENERGY	1.978103	1.504444	1.314840	0.1964
LENSQ	-0.136047	0.113961	-1.193799	0.2400
RSEMI	1.100024	0.477359	2.304393	0.0268
RLANDLES	0.000163	0.024887	0.006568	0.9948

R-squared	0.259656	Mean dependent var	1.156288
Adjusted R-squared	0.162243	S.D. dependent var	1.291541

U.S. Defense Expenditures Predicted by U.S. Gross National Product (1967-88)
(data in billions of current dollars)

Source	SS	df	MS
Model	124814.771	1	124814.771
Residual	7043.38113	20	352.169057
Total	131858.152	21	6278.95964

Number of obs =	22
F(1, 20) =	354.42
Prob > F =	0.0000
R-squared =	0.9466
Adj R-squared =	0.9439
Root MSE =	18.766

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.05967	.0031696	18.83	0.000	.0530584	.0662815
_cons	1.283966	8.417631	0.15	0.880	-16.2749	18.84284

Durbin-Watson Statistic = .2038266

Year	Residual	Actual	Predicted
1967	18.70228	67.4	48.69772
1968	25.33236	77.3	51.96764
1969	21.0528	77.8	56.7472
1970	16.73083	77.1	60.36917
1971	10.21051	74.5	64.28949
1972	5.01059	75.1	70.08941
1973	-4.54506	73.2	77.74506
1974	-8.20648	77.6	85.80648
1975	-7.23149	84.9	92.13149
1976	-14.71551	87.9	102.6155
1977	-21.02602	95.6	116.626
1978	-27.87521	103.0	130.8752
1979	-32.34412	115.0	147.3441
1980	-27.83859	132.8	160.6386
1981	-23.38235	156.1	179.4824
1982	-5.69398	182.9	188.594
1983	10.99837	210.5	199.5016
1984	6.07107	227.4	221.3289
1985	16.57645	253.7	237.1236
1986	22.64186	273.4	250.7581
1987	16.36612	282.0	265.6339
1988	3.16558	290.4	287.2344

Adding USSR Defense Expenditures Lagged by 1 Year (i.e., 1987
USSR Defense Expenditures used to predict 1988 US Defense Expenditures)

Source	SS	df	MS
Model	123805.037	2	61902.5184
Residual	2422.97787	18	134.609882
Total	126228.015	20	6311.40073

Number of obs =	21
F(2, 18) =	459.87
Prob > F =	0.0000
R-squared =	0.9808
Adj R-squared =	0.9787
Root MSE =	11.602

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	-.0175707	.0140985	-1.25	0.229	-.0471905	.0120491
ussrlag	1.290704	.2303581	5.60	0.000	.8067398	1.774669
_cons	24.10784	7.20209	3.35	0.004	8.976811	39.23887

Durbin-Watson Statistic = .5087065

Adding the Impact of the Vietnam War (Dummy Variable: 1967-74=1, 1975-88=0)

Source	SS	df	MS
Model	124590.357	3	41530.1191
Residual	1637.65716	17	96.3327742
Total	126228.015	20	6311.40073

Number of obs =	21
F(3, 17) =	431.11
Prob > F =	0.0000
R-squared =	0.9870
Adj R-squared =	0.9847
Root MSE =	9.8149

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.0071875	.0147458	0.49	0.632	-.0239233	.0382983
ussrlag	.9854533	.2222733	4.43	0.000	.5164977	1.454409
viet	22.21231	7.779598	2.86	0.011	5.798789	38.62582
_cons	-4.547886	11.74089	-0.39	0.703	-29.31899	20.22322

Durbin-Watson Statistic = .7741421

Adding the Impact of Political Variables (Presidency: Democrat=1, Republican=0; Percentage Democratic of the Senate and House)

Source	SS	df	MS
Model	124750.068	6	20791.6779
Residual	1477.94703	14	105.567645
Total	126228.015	20	6311.40073

Number of obs =	21
F(6, 14) =	196.95
Prob > F =	0.0000
R-squared =	0.9883
Adj R-squared =	0.9833
Root MSE =	10.275

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.0021632	.0174994	0.12	0.903	-.0353693	.0396958
ussrlag	1.01405	.252033	4.02	0.001	.4734927	1.554607
viet	9.725759	15.20277	0.64	0.533	-22.88095	42.33246
pres	-5.1075	7.057711	-0.72	0.481	-20.24479	10.02979
sen	.1377639	.6580693	0.21	0.837	-1.273654	1.549182
house	-.9235618	1.191839	-0.77	0.451	-3.479801	1.632678
_cons	57.44264	75.79653	0.76	0.461	-105.1248	220.01

Durbin-Watson Statistic = .8743016

Adding the Impact of Economic Variables (percentage unemployed and the size Of the federal deficit in billions of current dollars)

Source	SS	df	MS
Model	125144.684	8	15643.0855
Residual	1083.33061	12	90.2775508
Total	126228.015	20	6311.40073

Number of obs =	21
F(8, 12) =	173.28
Prob > F =	0.0000
R-squared =	0.9914
Adj R-squared =	0.9857
Root MSE =	9.5014

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.0249606	.0195872	1.27	0.227	-.0177161	.0676374
ussrlag	.6009132	.3151735	1.91	0.081	-.0857909	1.287617
viet	11.48378	17.2858	0.66	0.519	-26.17875	49.1463
pres	-10.05844	7.737255	-1.30	0.218	-26.91647	6.799596
sen	.5962803	.8183113	0.73	0.480	-1.186667	2.379228
house	-1.480244	1.152328	-1.28	0.223	-3.990952	1.030464
unempl	-3.002536	3.869256	-0.78	0.453	-11.43292	5.427848
def	-6.225034	3.037116	-2.05	0.063	-12.84234	.3922738
_cons	64.43759	83.91661	0.77	0.457	-118.401	247.2762

Durbin-Watson Statistic = 1.383571

Prais-Winsten Method of Correcting for Autocorrelation

Source	SS	df	MS
Model	3465.63053	8	433.203816
Residual	534.135896	12	44.5113247
Total	3999.76642	20	199.988321

Number of obs =	21
F(8, 12) =	9.73
Prob > F =	0.0003
R-squared =	0.8665
Adj R-squared =	0.7774
Root MSE =	6.6717

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.02421	.0135681	1.78	0.100	-.0053522	.0537723
viet	12.3678	11.43973	1.08	0.301	-12.55723	37.29283
ussrlag	.6251681	.2209481	2.83	0.015	.1437634	1.106573
pres	-2.961623	5.363733	-0.55	0.591	-14.64819	8.724948
sen	.2927714	.5663805	0.52	0.615	-.9412658	1.526808
house	-.1303799	.7183462	-0.18	0.859	-1.695522	1.434762
unempl	3.035824	2.158965	1.41	0.185	-1.668157	7.739805
def	-.6825273	1.753636	-0.39	0.704	-4.503372	3.138317
_cons	-16.80063	59.56815	-0.28	0.783	-146.5885	112.9872
rho	.9179513					

Durbin-Watson statistic (original) 1.383571
Durbin-Watson statistic (transformed) 1.046861

Cochrane-Orcutt Method of Correcting for Autocorrelation

Source	SS	df	MS
Model	10266.1871	8	1283.27338
Residual	154.76074	11	14.0691582
Total	10420.9478	19	548.470937

Number of obs =	20
F(8, 11) =	91.21
Prob > F =	0.0000
R-squared =	0.9851
Adj R-squared =	0.9743
Root MSE =	3.7509

us	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gnp	.031925	.0079371	4.02	0.002	.0144555	.0493945
viet	18.48806	6.697726	2.76	0.019	3.746466	33.22966
ussrlag	.7194277	.129186	5.57	0.000	.4350913	1.003764
pres	-4.882793	3.093965	-1.58	0.143	-11.69256	1.926977
sen	-.0163461	.3370397	-0.05	0.962	-.7581656	.7254734
house	.7294	.4588161	1.59	0.140	-.2804474	1.739247
unempl	6.395923	1.452023	4.40	0.001	3.200041	9.591805
def	2.119251	1.195024	1.77	0.104	-.5109791	4.749481
_inter	-121.5448	37.85226	-3.21	0.008	-204.8571	-38.23253
rho	0.7811	0.0241	32.36	0.000	0.7305	0.8316

Durbin-Watson statistic (original) 1.383571
Durbin-Watson statistic (transformed) 2.054069

Notice the difference in the coefficient values, t statistics, R^{2s}, transformed Durbin-Watson values and number of observations between the Prais-Winsten and Cochrane-Orcutt methods.