POL 502
Lecture 12
April 28, 2005

This covers performing GLS in MINITAB as well as other time series problems.

Looking at the Presidential approval data set:

What is the effect of the economy on presidential approval?

We can begin with a graph of the two.

Put Approval on the Y axis and National prospections on the X axis.

How are these calculated?

National prospections is 100 + the % of people who say that the economy will be better in 12 months minus the % of people who say the economy will be worse.

Approval is the % of people who say that they approve of the job _____ is doing as president.

A better way to look at this relationship is to look at the series over time.

Stat → time series → time series plot

Put the variables under Y – they will both be on the y-axis while the x axis is time.

Go to frame, click multiple graphs and this will put the series on the same graph.

Click data/time stamp and put in variable to mark the X-axis.

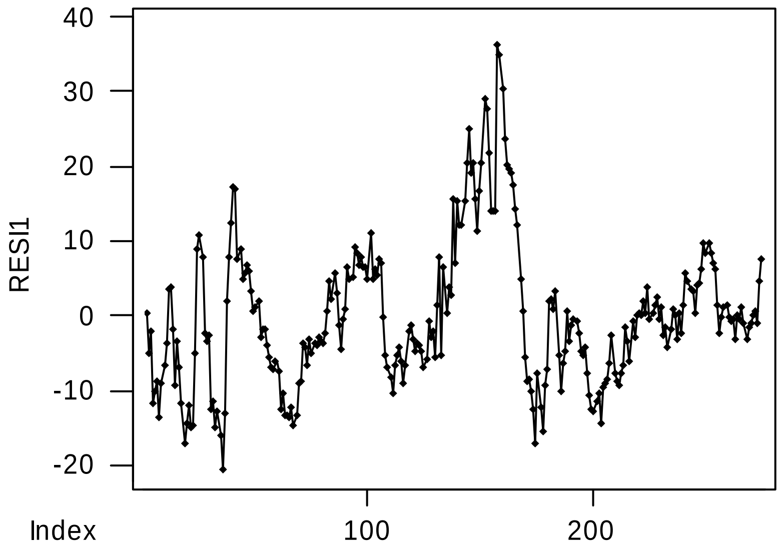
We can begin to look at the relationship between them using OLS regression:

```
The regression equation is
Approval = 37.8 + 0.146 NatPros

Predictor        Coef     SE Coef          T        P
Constant       37.756       2.083      18.12    0.000
NatPros       0.14568     0.01854       7.86    0.000

S = 9.576       R-Sq = 18.4%     R-Sq(adj) = 18.1%
```

But there is probably serial correlation going on.

So lets try that again, saving the residuals.

Now graph the residuals over time.



What should these look like?

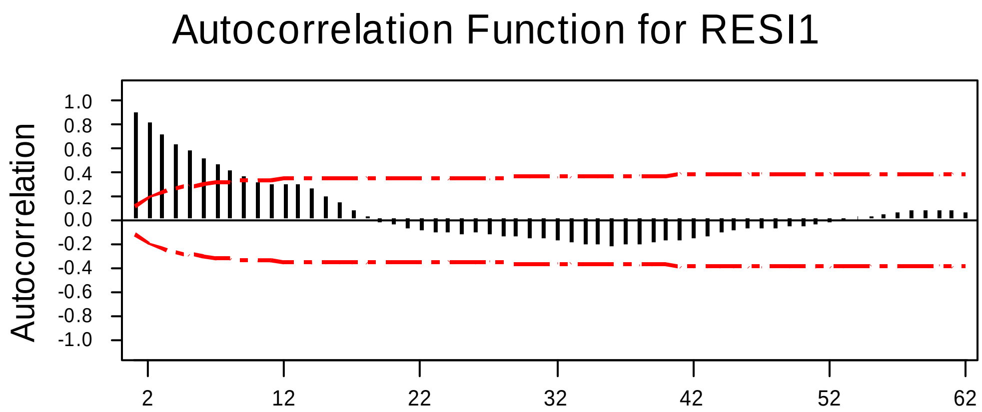
Random or "white noise."

Obviously the value of the error is dependent, at least in part, on its value in the previous period.

To further test for serial correlation, we can use the Durbin-Watson statistic.

Here we get a value of 0.20 which is very low and indicates high positive auto correlation.

The autocorrelation function tells us this as well.



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.90	14.97	226.62	16	0.14	0.76	1123.19	31	-0.16	-0.87	1177.27	46	-0.08	-0.40	1331.72	61	0.07	0.39	1349.18
2	0.81	8.32	410.73	17	0.07	0.40	1124.69	32	-0.18	-1.00	1187.91	47	-0.07	-0.36	1333.33	62	0.06	0.31	1350.48
3	0.71	5.96	553.22	18	0.02	0.13	1124.86	33	-0.20	-1.09	1200.66	48	-0.08	-0.39	1335.23				
4	0.64	4.75	667.52	19	-0.03	-0.18	1125.19	34	-0.21	-1.15	1214.88	49	-0.07	-0.34	1336.66				
5	0.57	3.98	761.08	20	-0.05	-0.28	1125.94	35	-0.22	-1.17	1230.01	50	-0.06	-0.32	1337.93				
6	0.52	3.41	837.80	21	-0.07	-0.38	1127.33	36	-0.22	-1.18	1245.60	51	-0.05	-0.25	1338.71				
7	0.47	2.94	899.81	22	-0.08	-0.47	1129.50	37	-0.22	-1.15	1260.45	52	-0.03	-0.14	1338.97				
8	0.42	2.55	949.64	23	-0.10	-0.56	1132.64	38	-0.20	-1.08	1273.95	53	-0.00	-0.01	1338.97				
9	0.36	2.16	987.23	24	-0.11	-0.62	1136.44	39	-0.19	-1.02	1285.96	54	0.01	0.06	1339.01				
10	0.31	1.81	1014.52	25	-0.12	-0.66	1140.85	40	-0.18	-0.97	1296.92	55	0.03	0.15	1339.33				
11	0.29	1.66	1038.29	26	-0.12	-0.64	1144.93	41	-0.18	-0.94	1307.33	56	0.04	0.20	1339.87				
12	0.29	1.66	1062.48	27	-0.13	-0.73	1150.27	42	-0.16	-0.85	1316.07	57	0.06	0.32	1341.23				
13	0.29	1.65	1086.88	28	-0.14	-0.75	1156.04	43	-0.14	-0.75	1322.94	58	0.07	0.37	1343.01				
14	0.26	1.45	1106.37	29	-0.14	-0.77	1162.07	44	-0.11	-0.58	1327.10	59	0.07	0.37	1344.81				
15	0.20	1.10	1117.69	30	-0.15	-0.84	1169.39	45	-0.09	-0.46	1329.71	60	0.08	0.42	1347.18				

We can see that there is a pattern to the error term – its value has influence far into the future.

What should we do?

One option is to use a lag of the dependent variable on the right-hand side.

$$y_t = a + \phi y_{t-1} + b_1 X_t + \varepsilon_t$$

$\phi = "phi"$ we call this the “autoregressive” parameter. Tells us how much of the previous value of y is ‘remembered’ in the present period.

We can create a lagged variable in MINITAB.

stat → time series → lag

Save a new series called Applag1 which is approval in the previous period.

Now look at the correlations between this and approval.

Again, a plot shows a strong relationship.

A regression shows:

```
The regression equation is
Approval = 4.29 + 0.920 Applag1

275 cases used 1 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant        4.291       1.302       3.30    0.001
Applag1       0.92037     0.02389      38.52    0.000

S = 4.187       R-Sq = 84.5%     R-Sq(adj) = 84.4%
```

With a DW of 1.81.

Here we are saying that approval at time *t* is .92 times the approval at time *t-1* plus some random error.

There is one missing case because we can't get the lagged value of the first observation.

Looking at these residuals over time looks to be much better – though there is still some autocorrelation left.

Check the autocorrelation function to see if these residuals have "time dependence."

Another option is differencing.

Here we move the lagged dependent variable over to the left hand side of the equation.

$$y_t - y_{t-1} = a + b_1 X_t + \varepsilon_t$$

We have changed the level of analysis from the value of *y* to changes in the value of *y*.

We will call this Δy_t.

$$y_t - y_{t-1} = \Delta y_t .$$

This is equivalent to assuming that $\phi = 1$ in the equation with the lagged dependent variable.

If we are to difference *y* we should also difference our independent variables – this makes interpretation much easier.

Another option is GLS using the Cochrane-Orcutt transformation.

Again we begin with:

$$y_t = a + b_1 X_t + \varepsilon_t \qquad (1)$$

But we transform *y* to get:

$y_t^* = y_t - \rho y_{t-1}$ (2)

Now we substitute (2) into (1) to give us:

$y_t^* = a + b_1 x_t + \varepsilon_t - \rho(a + b_1 x_{t-1} + \varepsilon_{t-1})$ (3)

Note how we substitute in for y_{t-1} using the fact that $y_{t-1} = a + b_1 x_{t-1} + \varepsilon_{t-1}$. (This makes sense given equation (1) above.)

Simplify (3) in a 2 steps:

$$y_t^* = a(1-\rho) + b_1(x_t - \rho x_{t-1}) + \varepsilon_t - \rho\varepsilon_{t-1}$$
$$y_t^* = a(1-\rho) + b_1(x_t - \rho x_{t-1}) + u_t$$

Where $u_t = \varepsilon_t - \rho\varepsilon_{t-1}$ and is a well behaved error term

Now let $a^* = a(1-\rho)$
and let $x_t^* = x_t - \rho x_{t-1}$

So that: $y_t^* = a^* + b_1 x_t^* + u_t$

Thus, the regression of y_t^* on x_t^* provides estimates of a^* and b_1 that are appropriate since the equation has an error term u_t that meets the assumptions.

We aren't interested in a^* necessarily but we can convert it back to *a* by dividing it by $1-\rho$.

So the Cochrane-Orcutt procedure is as follows:

1. determine an estimate of ρ. A good one to use is the entry for lag 1 in the ACF plot. We'll call this $\widehat{\rho}$.

2. form the transformed variables $y_t^* = y_t - \widehat{\rho} y_{t-1}$ and $x_t^* = x_t - \widehat{\rho} x_{t-1}$. We would do this for each additional independent variable.

3. Run the regression of y_t^* on the x_t^*'s.
The slope estimates are left alone.

The constant term estimate is adjusted by $\hat{a} = \dfrac{\widehat{a^*}}{(1-\widehat{\rho})}$.

This is just a computational trick that allows OLS to deal with auto correlation.

But the transformed values of the variables don't mean anything in and of themselves.

So, lets try this in MINITAB.

Step 1. Get an estimate of ρ .

stat → regression (save residuals) of approval on natpros.

stat → time series → autocorrelation → use residuals

Step 2. Create the transformed variables.

Create lagged *y*: stat → time series → lag (series: approval, storelags in: applag, lag: 1)

Create lagged *x*: stat → time series → lag (series: natpros, storelags in: nplag, lag: 1)

Create y_t^* : calc → calculator (store results in: ystar, expression: approval - .9*applag)

Create x_t^* : calc → calculator (store results in: xstar, expression: natpros - .9*nplag)

If we have any more independent variables we do the same thing to them.

Step 3. Run the GLS regression.

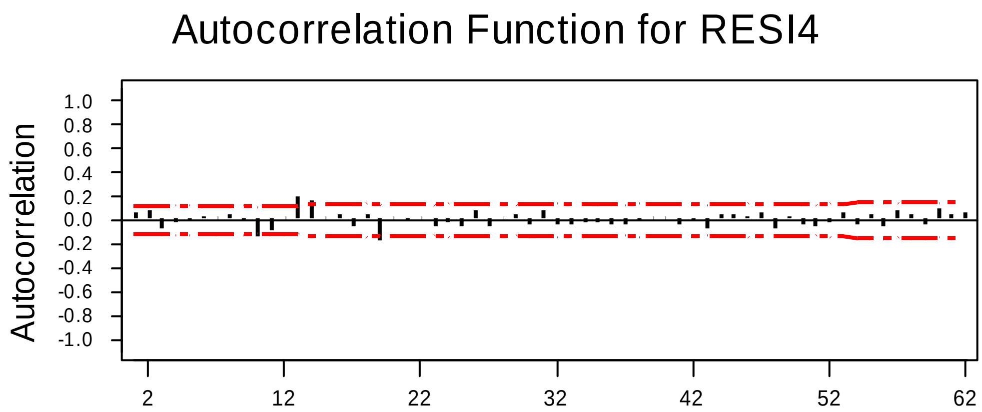
stat → regression (ystar is dependent variable and xstar is independent variable)

```
The regression equation is
ystar = 4.43 + 0.0878 xstar

275 cases used 1 cases contain missing values

Predictor        Coef     SE Coef          T         P
Constant       4.4279      0.3581      12.37     0.000
xstar         0.08777     0.02391       3.67     0.000

S = 4.092       R-Sq = 4.7%      R-Sq(adj) = 4.4%
```

Durbin-Watson is 1.89 – very good.

What do the diagnostics tell us?

What do the remaining autocorrelations look like?



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.05	0.88	0.77	16	0.04	0.59	31.83	31	0.08	1.10	50.78	46	0.02	0.26	57.42	61	0.05	0.64	72.41
2	0.08	1.32	2.56	17	-0.06	-0.94	32.99	32	-0.04	-0.56	51.26	47	0.06	0.91	58.81	62	0.05	0.72	73.39
3	-0.08	-1.28	4.26	18	0.04	0.64	33.53	33	-0.04	-0.58	51.78	48	-0.07	-1.02	60.59				
4	-0.02	-0.36	4.40	19	-0.18	-2.66	42.95	34	-0.02	-0.25	51.87	49	0.03	0.45	60.95				
5	-0.02	-0.27	4.48	20	0.01	0.10	42.96	35	-0.03	-0.47	52.22	50	-0.04	-0.51	61.39				
6	0.03	0.49	4.73	21	-0.00	-0.06	42.97	36	-0.05	-0.69	52.96	51	-0.06	-0.79	62.49				
7	0.00	0.07	4.74	22	0.01	0.11	42.99	37	-0.05	-0.70	53.72	52	-0.02	-0.24	62.59				
8	0.05	0.79	5.40	23	-0.06	-0.83	43.96	38	-0.01	-0.18	53.78	53	0.06	0.90	64.03				
9	-0.00	-0.05	5.40	24	-0.02	-0.28	44.08	39	0.00	0.04	53.78	54	-0.04	-0.54	64.55				
10	-0.14	-2.30	11.12	25	-0.06	-0.85	45.11	40	0.00	0.00	53.78	55	0.03	0.48	64.97				
11	-0.09	-1.50	13.67	26	0.07	1.07	46.78	41	-0.05	-0.67	54.50	56	-0.05	-0.73	65.93				
12	0.01	0.21	13.72	27	-0.06	-0.86	47.87	42	-0.00	-0.07	54.51	57	0.08	1.05	67.95				
13	0.19	2.95	23.78	28	0.00	0.06	47.87	43	-0.07	-1.03	56.24	58	0.04	0.55	68.52				
14	0.16	2.48	31.37	29	0.04	0.63	48.47	44	0.04	0.56	56.77	59	-0.04	-0.48	68.96				
15	0.01	0.08	31.38	30	-0.04	-0.57	48.95	45	0.04	0.57	57.31	60	0.09	1.20	71.64				

With no autocorrelation our estimate of $\widehat{\beta}$ is reliable and BLUE.

Distributed Lag Model

We can also use lags of independent variables.

$$y_t = a + b_1 X_t + b_2 X_{t-1} + b_3 X_{t-2} + \varepsilon_t$$

Effect of X on Y is spread over 3 periods or some number of finite periods.

Total effect of X = $b_1 + b_2 + b_3$

or: $b = \sum_{i=1}^{k} b_i$

$\dfrac{b_i}{\sum_{i=1}^{k} b_i}$ gives us the proportion of the affect appearing at time i.

Problems with this model:

1. With several X's and several lag we are estimating many parameters. We lose degrees of freedom and patience!

2. Problem of multicollinearity.

This will inflate standard errors.

Are there other variables that we could add to the model?

We can try unemployment contemporaneously, at lags of 1,2, and 3 months.