POL 502
Lecture 11
Prof. Matthew Lebo
April 21st, 2005

Serial Correlation and Heteroskedasticity

Serial correlation, also called autocorrelation, violates the Classical Assumption that observations of the error term are uncorrelated with each other.

Can exist in any study where the order of the observations has some meaning.

Can occur when the data are spatially related. For example, states that are next to each other may be more closely related than states that are far apart.

This violates the assumption that our cases are randomly chosen and don't affect each other.

Occurs most often in time series data, that is variables measured over time such as monthly measures of presidential approval and the economy.

If we have serial correlation it means that the error term from one time period depends in some systematic way on the error term from other time periods.

Time series is quite different from the OLS methods we have been learning.

It requires some different techniques and we must always be conscious of the possibility of serial correlation.

Traditional Time Series Analysis

Researchers using time series data in political science typically have used many of the same regression techniques as are employed to analyze cross-sectional data.

We can call this the *Econometric Approach*

Mostly single equation models such as the following:

$$Y_t = \beta_0 + \sum \beta_{1-k} X_{1-k,t-i} + \varepsilon_t$$

Where: Y_t is the dependent variable at time *t*.
X_{t-i} = 1 to *k* independent variables at time *t-i*.
β_0 = constant.
β_{1-k} = parameters associated with variables X_{1-k}

ε_t = stochastic error term $\sim N(0,\sigma^2)$

For a model such as above OLS is employed to estimate the parameters.

The effects of the X's may be specified to occur simultaneously (i.e. at time *t*, or with a lag *i*).

Inferences about the parameters are made using *t*-ratios.

When doing diagnostics on such regression models, particular attention is given to the possibility that the errors are correlated.

Assumption 4 in the text tells us that the errors are uncorrelated with each other.

That is: $E(r_{\varepsilon_i \varepsilon_j}) = 0$ for $i \neq j$.

We expect the correlations between time points to be zero regardless of the distance between them.

If each observation of the error term is correlated to the previous correlation we call this "first-order serial correlation."

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

Where ε is the error term.
ρ (rho) is the parameter that tells us how much the current value of ε depends on the previous value – this is a kind of *autoregressive* parameter.

u_t is an error term that meets the classical assumptions.

This is the most common type of serial correlation. We can have second order if the current observed error depends on the error of two periods ago, third-order and so on.

The magnitude of ρ tells us the strength of serial correlation.

If ρ = 0 there is no serial correlation.

As the absolute value of ρ approaches 1, the previous value of the error term, ε_{t-1} has a greater effect on the present value.

It is not reasonable for the absolute value of rho to be larger than 1, thus: $-1 \leq \rho \leq 1$.

The sign of ρ tells us the nature of serial correlation in an equation.

It may be positive or negative.

If its positive it means that a high error term in one period will lead to a higher one in the next period.

A negative value of ρ means that the error term has a tendency to switch signs from positive to negative and back again in consecutive observation. This is far less common than positive serial correlation.

Consequences of Serial Correlation

Pure serial correlation (that is, serial correlation in an otherwise correctly specified equation) does not cause bias in the coefficient estimates.

Serial correlation increase the variances of the $\widehat{\beta}$ distributions , that is, it creates inefficiency.

Standard errors are biased downwards leading to a higher possibility for Type I errors:

$t = \frac{\widehat{\beta}}{SE_{\widehat{\beta}}}$ gets larger and we are more likely to falsely reject the null hypothesis.

See Chapter 9 in the text for a more detailed discussion of these consequences.

How can we look for serial correlation?

Looking at the residuals of our series is a start.

Looking at the correlations in what is called an auto-correlation function (ACF) is best.

This will show us the correlations between time points separated by one period, 2 periods, and so on.

In MINITAB, “autocorrelation” is an option under the “time series” tab under stats that will create an ACF.

We can also use a partial auto-correlation functions (PACF) which tells us the same thing but holds the intervening correlations constant.

We can also find evidence of autocorrelation in the Durbin-Watson test statistic.

This is a test for 1^{st} order autocorrelation.

$$d = \frac{\sum(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum \hat{\varepsilon}_t^2} = 2(1-\rho)$$

where: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ ρ = Rho.

That is, ρ is the autoregressive parameter for the error term.

When ρ =0, d=2 and no autocorrelation exists.

Anything close to 2 indicates that autocorrelation is not a problem. The farther we get from 2, the more we should worry.

If there is perfect positive autocorrelation ρ =1 and d=0.
If there is perfect negative autocorrelation ρ =-1 and d=4.

Which makes one ask, if Durbin was so smart why can't 0 be neutral?

Values close to 0 or 4 indicate problems of 1st order autocorrelation.

You can use these numbers as a guide or find a table of critical values the test statistic in a Durbin's table.

So one way to think about autocorrelation is to think about ρ in the following:

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

Where ρ captures the relationship between temporally adjacent errors (that just means errors next to each other in time), and v_t is a well behaved error term, $\sim N(0,\sigma^2)$.

This relationship between the errors is treated as a nuisance to be corrected.

Correction employed is usually Generalized Least Squares.

Generalized Least Squares (GLS)

Involves multiplying both sides of the model by the "quasi-differencing" operator $(1-\rho L)$.

Suppose we perform a Durbin-Watson test on:

$y_t = a + bX_t + \varepsilon_t$ Call this Equation 1.

and conclude that the error process is 1st order autoregressive.

What do we do to fix this?

Assume first that we know ρ.

In GLS we do some kind of transformation.

Here, lag equation back 1 period and multiply by ρ.

Remember that $-1 \le \rho \le 1$.

$$\rho y_{t-1} = \rho(a + bX_{t-1} + \varepsilon_{t-1})$$
$$\rho y_{t-1} = \rho a + \rho bX_{t-1} + \rho\varepsilon_{t-1}$$

Call this Equation 2.

Create Equation 3:

Equation 3 = Equation 2 – Equation 1.

$$y_t - \rho y_{t-1} = (a - \rho a) + (bX_t - \rho bX_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

or:

$$y_t - \rho y_{t-1} = a(1-\rho) + b(X_t - \rho X_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

The last part is the most useful part because it equals v_t.

Call $(\varepsilon_t - \rho\varepsilon_{t-1})$ v_t which is now a well behaved error term.

Since we know what ρ, X, and Y are we can transform the variables and calculate a and *b*.

This is called the "**Cochrane Orcutt Transformation**."

The problem is that we don't usually know ρ.

So must estimate it.

This is called: *Pseudo GLS* or *Feasible GLS.*

Start with the original equation.

Save residuals.

Regress the residuals on itself lagged one period:

$$\hat{\varepsilon}_t = \hat{\rho}\hat{\varepsilon}_{t-1} + v_t$$

This gives an estimate of ρ.

We can transform Y and X and go back to the beginning.

This is one approach to time series analysis and handle the problem of autocorrelation.

Another is to include a "lagged endogenous variable." Endogenous = Dependent.

Lagged Endogenous (Dependent) Variables

If serial correlation exists, OLS gives biased but consistent estimates – as sample grows, the value of the parameter will converge on its natural value.

Why?

$$y_t = a + b_1 y_{t-1} + \varepsilon_t \qquad (1)$$

where, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ (2)

Note that *y* appears on both sides of equation (1) which becomes problematic.

Both ρ and *b* are bounded between -1 and +1.