OLS Assumptions about Error Variance and Covariance

- For OLS estimators to be BLUE,
  - $E(u_i)^2 = \sigma^2$ (Homoscedasticity)
  - $E(u_iu_j) = 0$ (No autocorrelation)
Problem of non-constant error variances is known as HETEROSCEDASTICITY

Problem of non-zero error covariances is known as AUTOCORRELATION

\[ E(u_i u_j) \neq 0 \text{ where } i \neq j \]

These are different problems and generally occur with different types of data

Implications for OLS are the same
Graphical Illustration of Autocorrelation

- Plotting the residuals over time will often show an oscillating pattern

- Correlation of $u_t$ & $u_{t-1} = .85$
Non-autocorrelated Series

As compared to a non-autocorrelated model
Causes of Autocorrelation

- Autocorrelation: problem in time-series data
  - Why? Error at one point in time might be correlated with error at another point in time.

- Spatial Autocorrelation may also occur: problem with cross-sectional data
  - Error is correlated across units
Other Causes of Autocorrelation

- Nonstationarity
  - Stationary: a time series is stationary if its characteristics (mean, variance and covariance) do not change over time.
    - i.e. no time trends
  - With time series data, the characteristics might fluctuate, thus the error term will also fluctuate and exhibit autocorrelation
Consequences of Autocorrelation

- Estimation in the presence of autocorrelated errors is the same as for heteroscedasticity
  - OLS is unbiased
  - OLS is not BEST: Inefficient
  - Estimate of $\text{var}(\mathbf{b})$ will be incorrect and therefore t-test will be wrong!
What is *Generalized* Least Squares (GLS)?

- Solution to both heteroskedasticity and autocorrelation is GLS
- GLS is like OLS, but we provide the estimator with information about the variance and covariance of the errors
- In practice the nature of this information will differ – specific applications of GLS will differ for heteroskedasticity and autocorrelation
From OLS to GLS

- Thus we need to define a matrix of information $\Omega$ or to define a new matrix $W$ in order to get the appropriate weight for the X’s and Y’s.
- The $\Omega$ matrix summarizes the pattern of variances and covariances among the errors.
From OLS to GLS

- In the case of heteroskedasticity, we give information in $\Omega$ about the variance of the errors.
- In the case of autocorrelation, we give information in $\Omega$ about the covariance of the errors.
What IS GLS?

- Conceptually what GLS is doing is weighting the data
- Notice we are multiplying X and Y by weights, which are different for the case of heteroskedasticity and autocorrelation.
- We weight the data to counterbalance the variance and covariance of the errors
Patterns of Autocorrelation

- Autocorrelation can be across one term:
  \[ u_t = \rho u_{t-1} + \varepsilon_t \]
  where \(-1 < \rho < 1\) (\(\rho\) (rho)=the coefficient of autocovariance) and \(\varepsilon\) is the stochastic disturbance term (satisfies properties of OLS-white noise error term).

- Or Autocorrelation can be a more complex function:
  \[ u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \varepsilon_t \]
Patterns of Autocorrelation

\[ u_t = \rho u_{t-1} + \varepsilon_t \]

- A first-order autoregressive or AR(1) process is a VERY robust estimator of temporal autocorrelation problems.
- AR(1) is robust because most of correlation from previous errors will be transmitted through the impact of \( u_{t-1} \).
- One exception to this is seasonal or quarterly autocorrelation.
Patterns of Autocorrelation

- Second exception is spatial autocorrelation
- Difficult to know what pattern to adjust for with spatial autocorrelation
- For most time-series problems, AR(1) correction will be sufficient
  - At least captures most of the temporal dependence
Detecting Autocorrelation

- Common statistic for testing for AR(1) autocorrelation is the Durbin-Watson statistic

\[ d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{n} \hat{u}_t^2} \]

- Durbin-Watson is the ratio of the distance between the errors to their overall variance

- The null hypothesis is \( \rho = 0 \) or no autocorrelation
General Test of Autocorrelation: The Breusch-Godfrey (BG) Test

- More general than DW
  - Allows for lagged DV on right hand side, high-order autoregressive processes such as AR(2), moving average models

- Basic Process
  - Start with bivariate model
    \[ Y_t = B_1 + B_2 X_t + u_t \]
  - Assume error term \( u_t \), follows a \( p \)th-order autoregressive, AR(\( p \)), scheme
    \[ u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t \]
  - Null Hypothesis
    \[ H_0: \rho_1 = \rho_2 = \ldots = \rho_p = 0 \]
Illustration of Diagnostic Tests

- Let’s look at…Presidential Approval
- We have quarterly data from 1949-1985
- *First, tell STATA you have time-series data with the command:*
  - `tsset yearq (our variable for time)`
- Now we will run the following model:
  - `Approval= inflation + u`
Results from Regression

reg yes inflation

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 172</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7441.25469</td>
<td>1</td>
<td>7441.25469</td>
<td>F( 1, 170) = 53.41</td>
</tr>
<tr>
<td>Residual</td>
<td>23684.8804</td>
<td>170</td>
<td>139.322826</td>
<td>R-squared = 0.2391</td>
</tr>
<tr>
<td>Total</td>
<td>31126.1351</td>
<td>171</td>
<td>182.024182</td>
<td>Adj R-squared = 0.2346</td>
</tr>
</tbody>
</table>

| yes | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-----|-------|-----------|------|------|----------------------|
| inflation | -2.011885 | .2752906 | -7.31 | 0.000 | -2.555314 -1.468457 |
| _cons    | 63.49827  | 1.478642  | 42.94 | 0.000 | 60.5794  66.41713   |
Diagnostic Steps

- First, let’s look at graphical methods
  - After regress, type:
    - predict r, resid
  - To generate graph, type:
    - Scatter r yearq
Graphical Illustration: There appears to be a pattern
Try Durbin Watson Test

- `dwstat`

- Durbin-Watson d-statistic( 2, 172) = .4166369

- Since the DW statistic is close to 0, know we have positive autocorrelation

- We do not know if the process is AR(1), so we might also try the Breusch-Godfrey Test
Try Breusch-Godfrey Test

- bgodfrey, lags(1 2 3)

Breusch-Godfrey LM test for autocorrelation

<table>
<thead>
<tr>
<th>lags(p)</th>
<th>chi2</th>
<th>df</th>
<th>Prob &gt; chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>103.867</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>105.196</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>105.196</td>
<td>3</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

H0: no serial correlation

Can clearly reject null hypothesis of no autocorrelation, and it appears that all of the lags are significant, can keep trying other lags.
Theory vs. statistical tests?

- Theory can also tell us when to expect autocorrelation
- In my view, a better guide
What do we do when we find autocorrelation?

- Make sure that model is correctly specified
  - No Omitted Variable Bias, Correct Functional Form
- If pure autocorrelation, can use generalized least squares method
- If large sample, can use Newey-West method to obtain standard errors corrected for autocorrelation
Working in stata

- Tsset the data first
  - `tsset panel_var time_var`
- `xtreg` runs a GLS estimator
Fixed Effects vs. Random Effects

- RE exploits both over-time and cross-sectional variation

- FE exploits only changes within units
  - Eliminates many potential omitted variables

- Data in first-differences
Handling Time Trends

- If two variables have the same time trend then the two variables will be positively correlated

- Deal with time as will any omitted variable
  - Control for it
  - year as a variable for a linear time-trend
  - year, year squared, year cubed for nonlinear
  - year dummies for time-specific shocks (not trends)