<u>Lecture 5: Assumptions of the Classical Linear Regression Model:</u>

<u>The Gauss-Markov Theorem</u>

***The most technical, feared, and necessary lecture: and only the beginning.**

***When estimating slopes from randomly sampled data. When is OLS the best method to use?**

Part 1: When is OLS, minimizing the sum of squared residuals, the "BEST" way to estimate slopes?

Before we get into this, here are some preliminaries:

A) What is an estimator?

a) OLS: Minimize RSS
b) GLS: Minimize RSS but with some ways to correct for problems
c) Minimize sum of absolute values of residuals
d) Bayesian methods:
Suppose our prior information is that the slope is 2. Let us collect some information and use Bayes' rule to update this prior.
e) Maximum likelihood: choose the slopes that are the most likely to have generated our random sample of observations.
f) Just draw a line that looks good!

B) What is a Linear Estimator?

C) What is <u>best</u>? BEST: Minimum variance, minimum expected distance between our estimate and the TRUE slope.

D) What is unbiased?

OK, so when is OLS (Minimizing Sum of Squared Residuals) Best?

The Gauss Markov theorem Gauss 1821 Markov 1912

If the first 6 assumptions below are true, then OLS is BLUE (The best linear unbiased estimator)

If the all 7 assumptions below are true, the OLS is BUE (The best unbiased estimator of all)

I. The regression model is linear in the coefficients,
is correctly specified,
and has an additive error term

linear in the coefficients

YES $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

YES $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + \varepsilon_i$

YES $y_i = \beta_0 + \beta_1 \ln(x_{1i}) + \beta_2 \frac{1}{x_{2i}} + \beta_3 10^{x_3} + \varepsilon_i$

YES $\ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

NO $y_i = \beta_0 + e^{\beta_1} x_{1i} + \beta_2{}^3 x_{2i} + \varepsilon_i$

NO $y_i = \beta_0 x_{1i}^{\beta_1} x_{2i}^{\beta_2} + \varepsilon_i$ Cobb-Douglas

YES $\ln(y_i) = \ln \beta_0 + \beta_1 \ln(x_{1i}) + \beta_2 \ln(x_{2i}) + \varepsilon_i$

is correctly specified

1. **Correct functional form**
2. **Correct set of variables**

and has an additive error term

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

MPG=32-0.06HP+1*Domestic+e_i

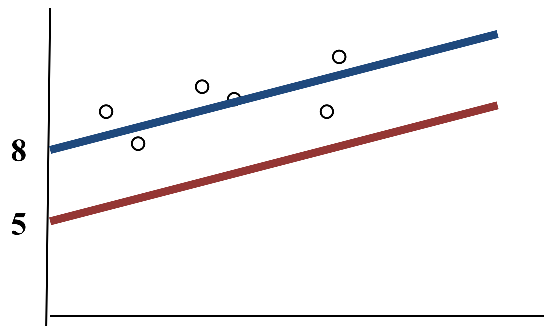
MPG=32-60+1=29MPG

Multiplicative Error Term?

$y_i = [\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}]\varepsilon_i$

MPG=[32-0.06HP+1*Domestic]*e_i

II. The error term has a zero population mean. (We usually don't worry about it, but there should be no theoretical reason to think otherwise.)



$$y_i = 5 + \beta_1 x_{1i} + \varepsilon_i (avg = 3)$$
$$y_i = 8 + \beta_1 x_{1i} + \varepsilon_i (avg = 0)$$

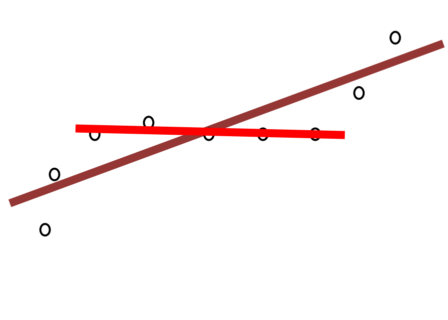
III. All explanatory variables are uncorrelated with the error term.

Simultaneity

$$Crime_i = \beta_0 + \beta_1 Income_i + \beta_2 Police_i + \varepsilon_i$$
$$Crime_i = \beta_0 + \beta_1 Income_i + \beta_2 Police_i + 50_i$$

IV. Observations of the error term are uncorrelated with each other (no serial correlation).

GDP



V. The error term has a constant variance, homoskedasticity (no heteroskedasticity).

VI. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity)

Correlation= +1 = -1

Electricity = B0 +B1*Celsius +B2*F

F = (9/5)C + 32

Domestic =1 =0 import
Import =1 =0 domestic

N S E W

N=0 N=1-S-E-W
S=0
E=0
W=1

(severe imperfect multicollinearity is also a problem, but not really a violation of this assumption).

VII. The error term is normally distributed (Optional for BLUE, but if true OLS is the BEST method of all possible methods!)

I. The regression model is linear in the coefficients,

is correctly specified (Ch. 6 & 7),

and has an additive error term

II. The error term has a zero population mean. (We usually don't worry about it, but there should be no theoretical reason to think otherwise.)

This means that the expected value of ε_i is zero ($E[\varepsilon_i]=0$).
Because Y_i-hat = $Y_i + \varepsilon_i$, this means that: $E[Y_i\text{-hat}] = Y_i + 0$
So, the expected value of Y_i-hat is equal to the actual value of Y_i.
Practically though, if the true mean of the error term is not zero, it will be absorbed by the y intercept.

III. All explanatory variables are uncorrelated with the error term.
This means that when an error term is larger, it cannot cause an explanatory variable to increase or decrease. Usually from *simultaneous equations*(Chapter 14). In Macroeconomics:
$C = B_0 + B_1Y + e$, but we know that Y=C+I+G
If e was 10 this year, then:
1) What happens to C?
2) Then how does this affect Y?
3) Then how does this affect C?
4) Then how does this affect Y?………
So we cannot tell exactly how Y affects C with OLS. But, it can be done other ways…

IV. Observations of the error term are uncorrelated with each other (no serial correlation).
(Chapter 9) If you're looking at time series data (data collected from the same source in a number of different periods) the error term (Y_i - Y_i-hat) in one period should not have any relation to the error term from the previous period. In cross-sectional data, a pattern in error means wrong functional form!

V. The error term has a constant variance, homoskedasticity (no heteroskedasticity).
This means that the errors aren't more spread out for some of the observations than for others. (Chapter 10)

VI. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity (or severe imperfect multicollinearity)).
Generate a matrix of correlation coefficients for all the explanatory variables and the dependent variable. The standard errors of the OLS estimates of the parameters of the collinear variables are very large. These high variances arise because in the presence of multicollinearity the OLS estimating procedure is not given enough independent variation in a variable to calculate the effect it has on the dependent variable with any confidence!

VII. The error term is normally distributed (Optional for BLUE, but if true OLS is the BEST method of all possible methods!)
This is important when generating confidence intervals and doing hypothesis testing in small samples but is usually less important as sample sizes increases due to the central limit theorem.