OLS Diagnostics

1. Heterosckedasticity

- $(u_i) = E(u_i^2|Xi) = \sigma^2$; unlikely in cross-section $var(u_i) = \sigma_i^2$
- Effect: Unbiased, consistent OLS, but less efficient [not BLUE]; unreliable statistical inference (estimated t values)
- Test:
  - Plot residuals against Yhat
  - Breusch-pagan, white test [H0 : homoscedastic error variance]
- Solution: Heterosckedasticity-consistent standard errors

2. Autocorrelation

- covariance E(UiUj) = 0 for I ≠ j; unlikely with time series
- Effect: Unbiased, consistent OLS, no longer efficient, unreliable se
- Test:
  - Plot residuals chronologically
  - Durbin Watson, Breusch-Godfrey LM tests
- Solution: first difference; HAC; introduce lags

3. Specification errors

- no specification bias – model 'correctly' specified
  - Underfitted models (excluding important variables): ramsey RESET test/Lagrange multiplier LM test
  - Overfitted models (including irrelevant variables): exclusion restriction F-test
  - Incorrect functional form
  - Measurement errors

4. Multicollinearity

- This is a problem which occurs when there is an exact or near exact linear relationship among Xs

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i$

- e.g., Perfect collinearity: $X_{1i} = 2 - 3X_{2i}$
- e.g., Perfect collinearity: income measured in dollars and cents, dummy variable trap
- However, exact linear relationships among regressors is a rarity
- e.g., Imperfect collinearity: $X_{1i} = 2 - 3X_{2i} + v_i$
- Imperfect collinearity is more common than perfect collinearity

❑ Effects of multicollinearity:

- OLS estimators are still BLUE
- High R2 but several insignificant coefficients
- Regression coefficients are sensitive to small changes in data, especially with relatively small samples
- Difficult to isolate the impact of each variable separately on Y.

❑ Testing for collinearity:- there is no unique test for multicollinearity

- High pairwise correlations among explanatory variables
- High partial coefficients
- Significant F-test for auxiliary regressions
- High variance inflation factor

❑Solution for multicollinearity:

- Dropping variables may attenuate the problem (but we have to be careful of omitted variable bias)
- Principal component analysis
  - ➢Artificial variables which are orthogonal to one another
  - ➢Principal components become regressors in the model
  - ➢Interpretation of coefficients is not straightforward