MGMG 522 : Session #5
Multicollinearity

(Ch. 8)

5-1

Multicollinearity

- Multicollinearity is a violation of the classical assumption #6.
- There are two kinds of multicollinearity: perfect and imperfect multicollinearity.
- Perfect: if any X can be written as a linear combination of other X's.
- Imperfect: if any X is close to being represented by a linear function of other X's.

5-2

Multicollinearity Problem

- If perfect multicollinearity exists in the sample, OLS regression will not be able to distinguish the effect of one X, holding constant the effects of other X's.
- As a result, OLS will not be able to produce the coefficient estimates.
- This problem can be easily detected and solved.
- OLS can produce the coefficient estimates in the imperfect multicollinearity case, making this kind of problem harder to detect and solve.
- Example: The model is $Y = \beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$
  - Perfect case: $X_1 = \alpha_0+\alpha_1X_2$
  - Imperfect case: $X_1 = \alpha_0+\alpha_1X_2+u$

Dominant Variable

- A special case of multicollinearity problem between Y and X.
- If one of the X's is highly correlated with Y, that X variable is a dominant variable.
- That X variable will make other X's unimportant in determining the values of Y.
- For example, if your Y variable is the number of cars manufactured and you use the number of car engines as your X variable, you will likely have a dominant variable case.
- Do not confuse a dominant variable with a highly significant independent variable.

Imperfect Multicollinearity

- Perfect multicollinearity is rare.
- Therefore, when we refer to multicollinearity, we mean the imperfect multicollinearity case.
- Multicollinearity is a sample as well as theoretical phenomenon.

5-5

Consequences of Multicollinearity

1. Coefficient estimates will still be unbiased.
2. But, the variances and standard errors of the estimates will increase, $SE(\hat{\beta}_1) = \sqrt{\frac{\left(\sum e^2\right)/(n-3)}{\sum\left(X_1 - \overline{X}_1\right)^2\left(1 - r_{12}^2\right)}}$
3. However, OLS estimator still is BLUE.
4. t-values will be lower (because S.E. ↑), while Adj-R^2 is largely unaffected.
5. Coefficient estimates are sensitive to changes in specification.

5-6

Detecting Multicollinearity

- Multicollinearity often exists in any sample.
- Therefore, the question is not whether it exist, but rather how much is the multicollinearity problem?
- Remember, degree of multicollinearity problem can vary from sample to sample.
- There are no formal multicollinearity tests. But the following two tools may be somewhat helpful in detecting the multicollinearity problem.

5-7

1. Simple Correlation Test

- If r, the simple correlation coefficient, between two variables is high, this could indicate the existence of multicollinearity.
- How high?
  - Some suggest $|r|>0.8$
  - Some suggest t-test for significance.
- There is one problem: this test can only be performed on a pair of variables.

5-8

2. Variance Inflation Factors (VIF)

- Model: $Y=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k+\varepsilon$
- To find VIF of each X,

1. Regress each X on all other Xs and obtain a value of R^2.
2. Repeat this for all X's.
3. VIF for each X = $1/(1-R^2)$
4. If VIF > 5, multicollinearity problem is potentially severe.

Note: there are k VIFs in this example.

5-9

Remedies for Multicollinearity

1. Do nothing
2. Drop a redundant variable
3. Transform a variable
4. Increase the sample size

- Which remedy is most appropriate?
  - More often, doing nothing is the most appropriate course of action.

5-10