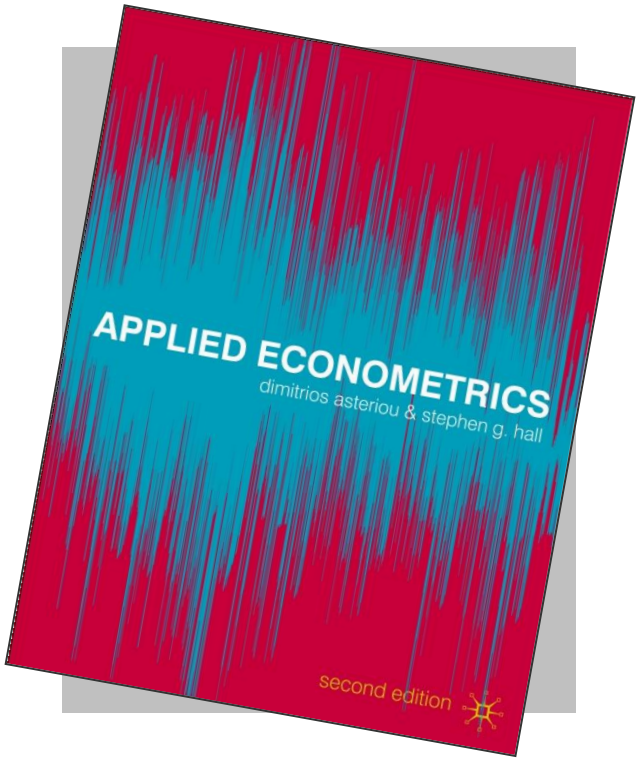
Applied Econometrics
Second edition

Dimitrios Asteriou and
Stephen G. Hall



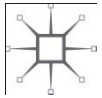
Chapter 9:
Dummy Variables

palgrave
macmillan

Applied Econometrics

Dummy Variables

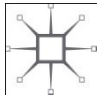
1. Nature of qualitative information
2. Use of dummy variables
3. Special cases of dummy variables
4. Dummy variables with multiple categories
5. Tests for structural stability



Applied Econometrics

Learning Objectives

1. Understand various forms of possible misspecification in the CLRM
2. Appreciate the importance and learn the consequences of omitting influential variables in the CLRM
3. Distinguish among wide range of functional forms and know meaning and interpretation of their coefficients
4. Understand importance of measurement errors in data
5. Perform misspecification tests using econometric software
6. Understand meaning of nested and non-nested models
7. Be familiar with concept of data mining and choose appropriate econometric model



Nature of Qualitative Information

Sometimes cannot obtain set of numerical values for all variables to use in a model.

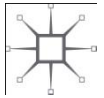
Because some variables cannot be quantified easily.

Examples:

(a) Gender may play a role in determining salary levels

(b) Different ethnic groups may follow different consumption patterns

(c) Educational levels can affect earnings from employment



Nature of Qualitative Information (2)

Easier to have dummies for cross-sectional variables, but sometimes we have for time series as well.

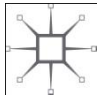
Examples:

(a) Changes in political regime may affect production

(b) War can impact on economic activities

(c) Certain days in week or certain months in year can have different effects on the fluctuation of stock prices

(d) Seasonal effects often observed in demand of various products



Applied Econometrics

Use of Dummy Variables

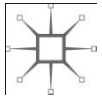
Consider following cross-sectional model:

$$Y_i=\beta_1+\beta_2 X_{2i}+ u_i$$

The constant term in this equation measures the mean value of Y_i when X_{2i} is equal to zero.

This model assumes that the constant will be the same for all the observations in our data set.

But what if we have two different subgroups (male and female, for example)?



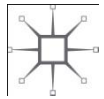
Use of Dummy Variables (2)

Question is how to quantify the information that comes from the difference in the two groups.

One solution is to create a dummy variable as follows:

$$D = \begin{cases} 1 & \text{for male} \\ 0 & \text{for female} \end{cases}$$

Note that the choice of which of the two different outcomes is to be assigned the value of 1 does not alter the results.



Use of Dummy Variables (3)

Entering this dummy in the equation we have the following model:

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3D_i+u_i$$

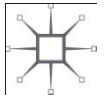
We have two cases:

$D_i=0$ $\quad Y_i=\beta_1+\beta_2X_{2i}+\beta_3(0)+u_i$

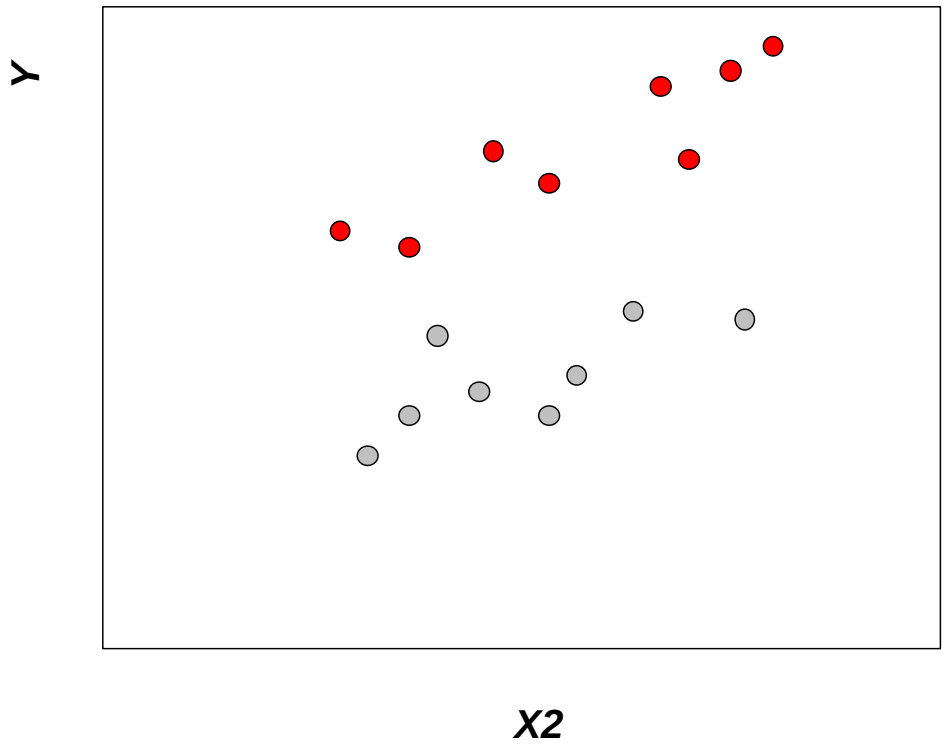
$$Y_i=\beta_1+\beta_2X_{2i}+u_i$$

$D_i=1$ $\quad Y_i=\beta_1+\beta_2X_{2i}+\beta_3(1)+u_i$

$$Y_i=(\beta_1+\beta_3)+\beta_2X_{2i}+u_i$$



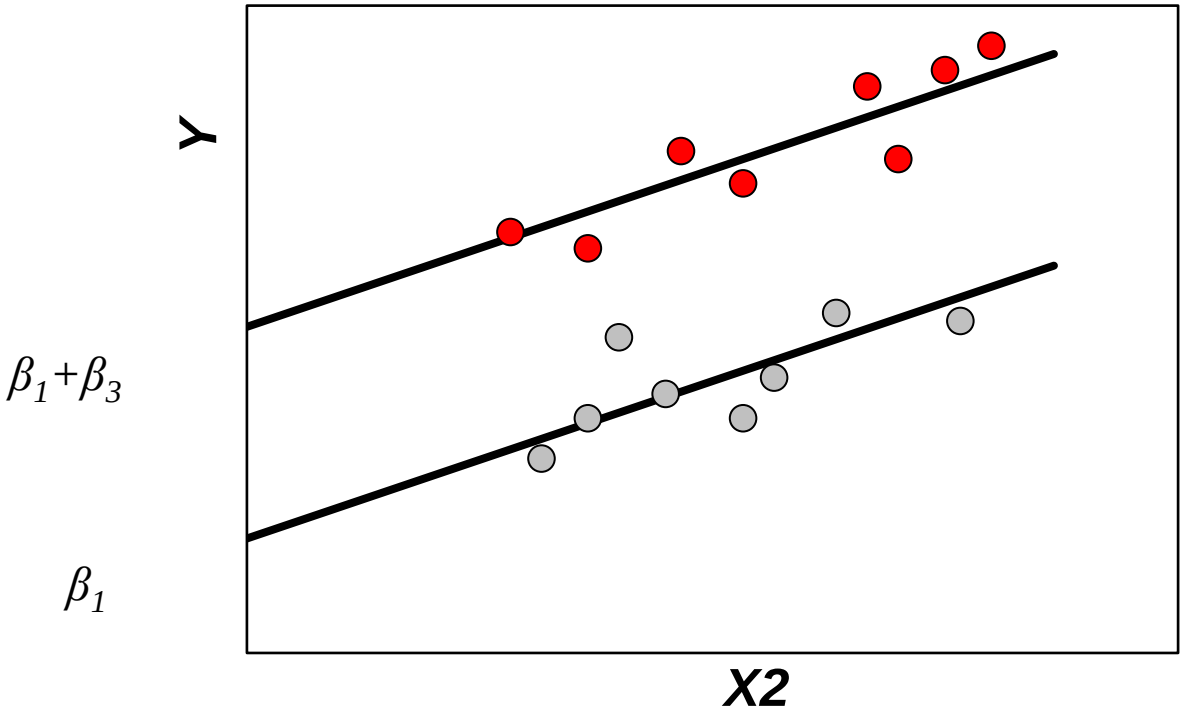
Intercept Dummy Variable



Red points: **Male**

Grey points: **Female**

Intercept Dummy Variable (2)



Slope Dummy Variable

Consider same case but now with the dummy affecting the slope

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3D_iX_{2i}+u_i$$

Two cases:

$D_i=0$

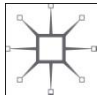
$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3(0)_iX_{2i}+u_i$$

$$Y_i=\beta_1+\beta_2X_{2i}+u_i$$

$D_i=1$

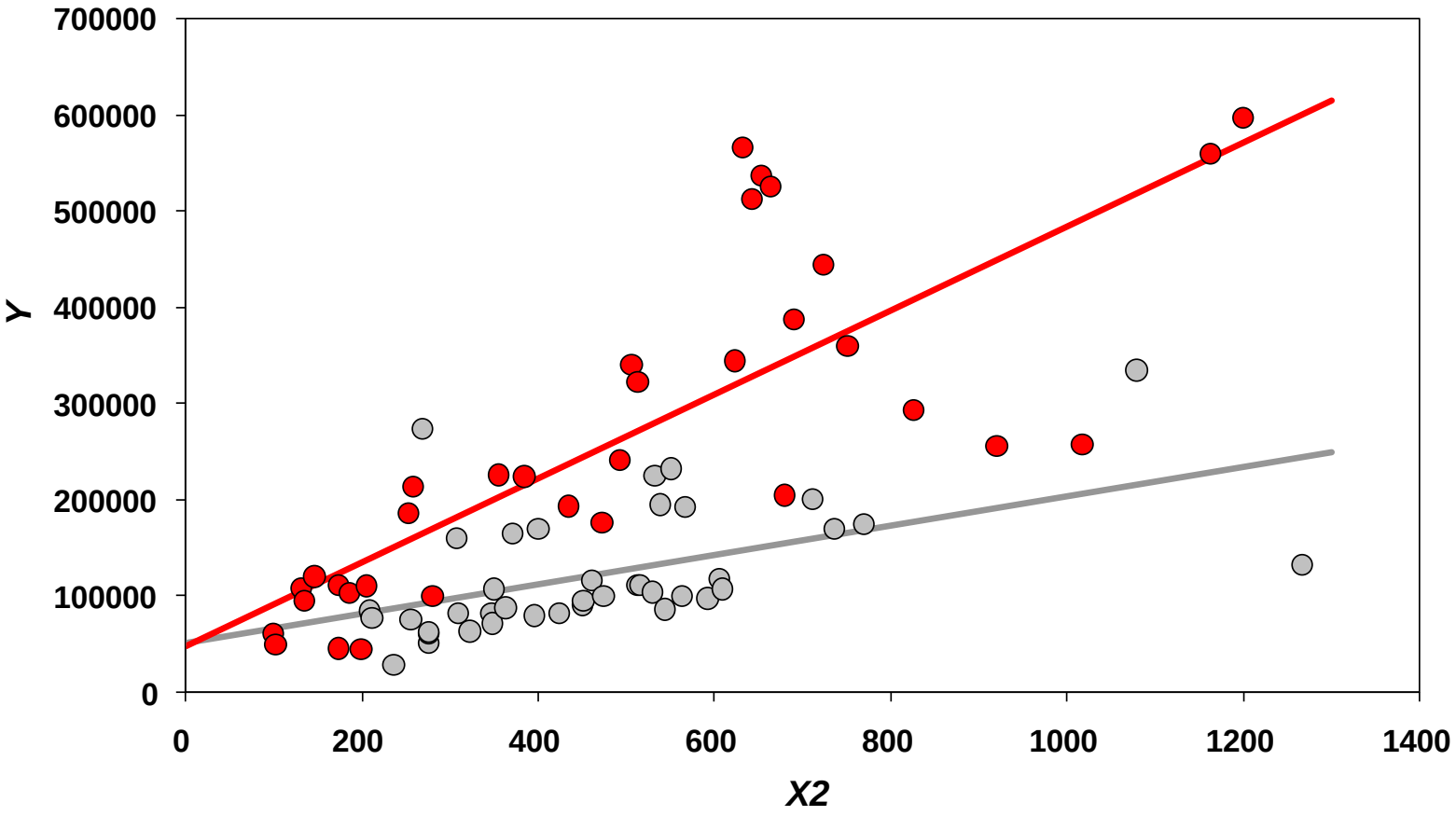
$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3(1)_iX_{2i}+u_i$$

$$Y_i=\beta_1+(\beta_2+\beta_3)X_{2i}+u_i$$

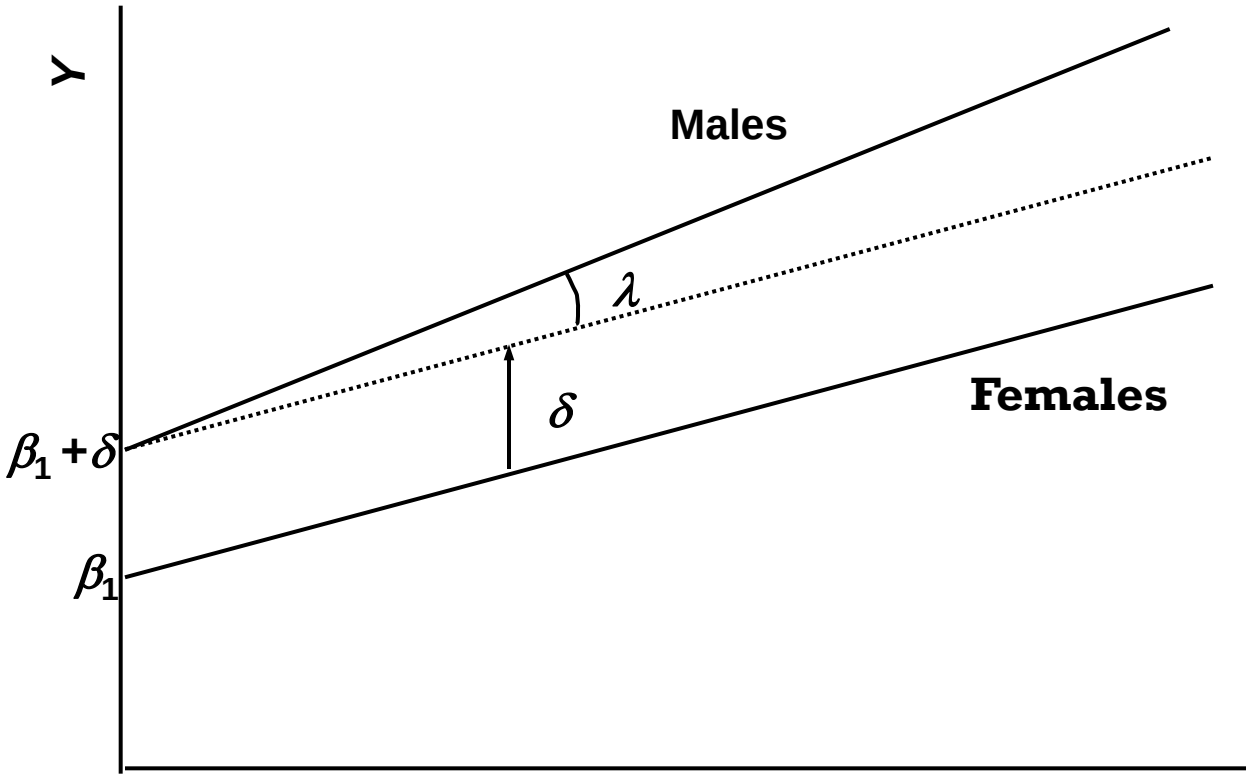


Applied Econometrics

Slope Dummy Variable (2)



Combined Effect



Applied Econometrics

Dummies with Multiple Categories

Consider the example of education:

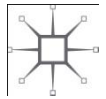
D1= 1 primary; 0 otherwise

D2= 1 if secondary; 0 otherwise

D3= 1 if tertiary; 0 otherwise

D4 = 1 if BSc; 0 otherwise

D5 = 1 if MSc; 0 otherwise



The Plug-in Solution

We estimate:

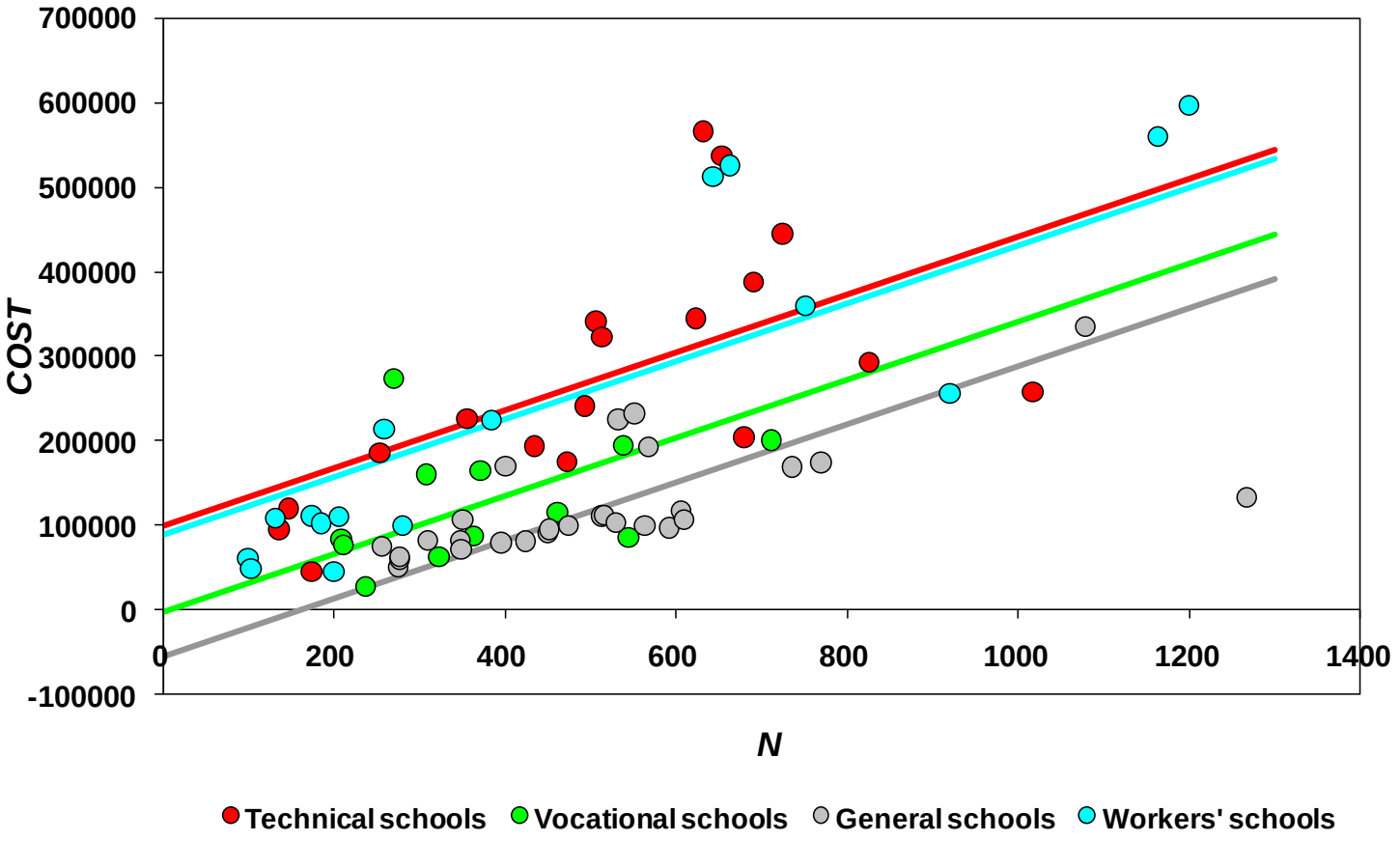
$$Y=\beta_1+\beta_2 X_2+ a_1 D_2+a_2 D_3+a_3 D_4+a_4 D_5+u$$

Note that one dummy (in this case D1) is excluded from the model in order to avoid dummy variable trap.

Consider various cases, i.e. D2=1, D3=D4=D5=0

Applied Econometrics

Multiple Categories



Applied Econometrics

Using more than One Dummy

GENDER (male; female)

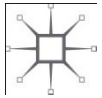
EDUC (primary, secondary, tertiary; BSc; MSc)

AGE (less than 30, 30 to 40; more than 40)

OCUP (unskilled, skilled, clerical, self-employed)

…and so on

The interpretation (although it seems more complicated) is the same as before



Seasonal Dummy Variables

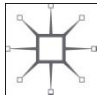
Depends on frequency of data:

Quarterly – 4 dummies – DQ1, DQ2, DQ3, DQ4

Monthly – 12 dummies – one for each month

Daily – 5 dummies – Dmon, Dtue, Dwed, etc.

Again, we either exclude one and include constant (always better) or if we use all we never include constant (dummy variable trap)



Chow Test for Structural Stability

Step 1: Estimate the basic regression equation

$$Y_i=\beta_1+\beta_2 X_{2i}+u_i$$

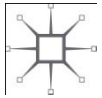
for three different data sets:

(a) whole sample, n;

(b) period before shock, n1; and

(c) period after shock, n2.

Step 2: Obtain the SSR for each of the three subsets and label them SSRN, SSR1 and SSR2.

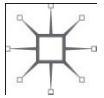


Chow Test for Structural Stability (2)

Step 3: Calculate the F-statistic

$$F = \frac{(SSR_N - [RSS_1 + RSS_2])/k}{(RSS_1 + RSS_2)/(n-2k)}$$

Step 4: If F-statistical bigger than F-critical *F(k,n-2k)* then reject the null that the parameters are stable for the whole data set.



Dummy Variable Test for Structural Stability

Dummy Variable Test much better because:

- Single equation used to provide set of estimated coefficients for two/more structures
- Only one degree of freedom is lost for every dummy used
- Larger sample used for estimation
- Provides information regarding the exact nature of the parameter instability