UN3412 Introduction to Econometrics

Recitation 6

Sean Hyland

Columbia University

Spring 2019

Outline

1. S&W Exercise 8.2
2. S&W Exercise 8.10
3. SW Exercise 9.6
4. S&W Additional Empirical Exercise 8.1
5. S&W Additional Empirical Exercise 9.1
6. Cobb-Douglas Production Function

Outline

1. S&W Exercise 8.2
2. S&W Exercise 8.10
3. SW Exercise 9.6
4. S&W Additional Empirical Exercise 8.1
5. S&W Additional Empirical Exercise 9.1
6. Cobb-Douglas Production Function

S&W Exercise 8.2 (i)

Suppose a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

(a) Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a house? Construct a 95% confidence interval for the percentage change in price.

(b) Comparing columns (1) and (2), is it better to use *Size* or ln(*Size*) to explain house prices?

(c) Using column (2), what is the estimated effect of a pool on price? (Make sure you get the units right.) Construct a 95% confidence interval for this effect.

S&W Exercise 8.2 (ii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (iii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (iv)

Suppose a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

(a) Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a house? Construct a 95% confidence interval for the percentage change in price.
An increase in size of 500 square feet, holding other factors constant, increases (ln) house prices by $500 \times 0.00042 = 0.21$, which is approximately a 21% increase in the house price. The 95% confidence interval for the percentage change in price is
$100 \times 500 \times (0.00042 \pm 1.96 \times 0.000038) = (17.276\%, 24.724\%)$.

(b) Comparing columns (1) and (2), is it better to use *Size* or ln(*Size*) to explain house prices?

(c) Using column (2), what is the estimated effect of a pool on price? (Make sure you get the units right.) Construct a 95% confidence interval for this effect.

S&W Exercise 8.2 (v)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (vi)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (vii)

Suppose a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

(a) Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a house? Construct a 95% confidence interval for the percentage change in price.

(b) Comparing columns (1) and (2), is it better to use *Size* or ln(*Size*) to explain house prices?
Since both (1) and (2) have the same dependent variable, we can look at the $\bar{R}^2$ to compare the goodness of fit of the two regressions. Regression (2) has higher $\bar{R}^2$ so it may be better to use ln(Size) to explain house prices, but the difference is pretty small. More intuitively, the log-log relationship seems more appropriate; increasing the size of a home by x square foot should have a different proportional effect on prices between small and large homes.

(c) Using column (2), what is the estimated effect of a pool on price? (Make sure you get the units right.) Construct a 95% confidence interval for this effect.

S&W Exercise 8.2 (viii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
$Pool \times View$					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (ix)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (x)

Suppose a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

(a) Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a house? Construct a 95% confidence interval for the percentage change in price.

(b) Comparing columns (1) and (2), is it better to use *Size* or ln(*Size*) to explain house prices?

(c) Using column (2), what is the estimated effect of a pool on price? (Make sure you get the units right.) Construct a 95% confidence interval for this effect. Adding a pool holding other factors constant, is expected to increase the (ln) house price by 0.071, which is an approximate 7.1% increase in house prices. The 95% confidence interval for the percentage change in price is $100 \times (0.071 \pm 1.96 \times 0.034) = (0.436, 13.764)$.

S&W Exercise 8.2 (xi)

(d) The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)

(e) Is the quadratic term $\ln(Size)^2$ important?

(f) Use the regression in column (5) to compute the expected change in price when a pool is added to a house that doesn't have a view. Repeat the exercise for a house that has a view. Is there a large difference? Is the difference statistically significant?

S&W Exercise 8.2 (xii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
$Pool \times View$					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xiii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	[illegible] (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xiv)

(d) The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)
With an additional bedroom and all other factors held constant, the (ln) house price is expected to increase by 0.0036, or approximately 0.36%. The effect is not statistically significant at the 5% level as the t-statistic is less than 1, which is less than the 5% critical value of 1.96. The reason the effect is so small is because the size of the home is being held constant in the regression, thus the addition of an extra bedroom must be accommodated through the reduction in the area of other rooms. This regression tells us that home buyers do not value the number of bedrooms independently of home size.

(e) Is the quadratic term $\ln(Size)^2$ important?

(f) Use the regression in column (5) to compute the expected change in price when a pool is added to a house that doesn't have a view. Repeat the exercise for a house that has a view. Is there a large difference? Is the difference statistically significant?

S&W Exercise 8.2 (xv)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xvi)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xvii)

(d) The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)

(e) Is the quadratic term $\ln(Size)^2$ important?
From the regression in column (4) we can see that the quadratic term on ln(Size) is not statistically important when the other set of regressors are included, as the t-statistic is approximately 0.06, thus well below any conventional critical value

(f) Use the regression in column (5) to compute the expected change in price when a pool is added to a house that doesn't have a view. Repeat the exercise for a house that has a view. Is there a large difference? Is the difference statistically significant?

S&W Exercise 8.2 (xviii)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
$Pool \times View$					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xix)

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
Pool × *View*					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

S&W Exercise 8.2 (xx)

(d) The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)

(e) Is the quadratic term $\ln(Size)^2$ important?

(f) Use the regression in column (5) to compute the expected change in price when a pool is added to a house that doesn't have a view. Repeat the exercise for a house that has a view. Is there a large difference? Is the difference statistically significant?

The value of a house without a view is expected to increase by (approximately) 7.11% when a swimming pool is added, all other factors are held constant.

The house price is expected to increase by $100\% \times (0.0711 + 0.00221) = 7.32\%$ when a swimming pool is added to a house with a view, holding other factors constant.

The difference in the expected change in price is reflected in the coefficient on the interaction term Pool*View: 0.0022. However, we note that this difference is not statistically significant at a 5% significance level as the t-statistic is $0.0022/0.10 = 0.022$, which is less than 1.96.

Outline

1. S&W Exercise 8.2
2. S&W Exercise 8.10
3. SW Exercise 9.6
4. S&W Additional Empirical Exercise 8.1
5. S&W Additional Empirical Exercise 9.1
6. Cobb-Douglas Production Function

S&W Exercise 8.10 (i)

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{i1} X_{2i} + u_i$$

Use Key Concept 8.1:

$$\Delta Y = f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2)$$

to show that

(a) $\Delta Y / \Delta X_1 = \beta_1 + \beta_3 X_2$ (effect of change in X_1, holding X_2 constant)

(b) $\Delta Y / \Delta X_2 = \beta_2 + \beta_3 X_1$ (effect of change in X_2, holding X_1 constant).

(c) If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$.

S&W Exercise 8.10 (ii)

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{i1} X_{2i} + u_i$$

Use Key Concept 8.1:

$$\Delta Y = f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2)$$

to show that

(a) $\Delta Y / \Delta X_1 = \beta_1 + \beta_3 X_2$ (effect of change in X_1, holding X_2 constant)

$$\begin{aligned}
\Delta Y =& f(X_1 + \Delta X_1, X_2) - f(X_1, X_2) \\
=& \left(\beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3 (X_1 + \Delta X_1) X_2 + u\right) \\
& - \left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u\right) \\
=& \beta_1 \Delta X_1 + \beta_3 \Delta X_1 X_2 \\
=& \left(\beta_1 + \beta_3 X_2\right) \Delta X_1
\end{aligned}$$

(b) $\Delta Y / \Delta X_2 = \beta_2 + \beta_3 X_1$ (effect of change in X_2, holding X_1 constant).

(c) If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$.

S&W Exercise 8.10 (iii)

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{i1} X_{2i} + u_i$$

Use Key Concept 8.1:

$$\Delta Y = f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2)$$

to show that

(a) $\Delta Y / \Delta X_1 = \beta_1 + \beta_3 X_2$ (effect of change in X_1, holding X_2 constant)
(b) $\Delta Y / \Delta X_2 = \beta_2 + \beta_3 X_1$ (effect of change in X_2, holding X_1 constant).

$$\begin{aligned}
\Delta Y =& f(X_1, X_2 + \Delta X_2) - f(X_1, X_2) \\
=& (\beta_0 + \beta_1 X_1 + \beta_2 (X_2 + \Delta X_2) + \beta_3 X_1 (X_2 + \Delta X_2) + u) \\
& - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u) \\
=& \beta_2 \Delta X_2 + \beta_3 X_1 \Delta X_2 \\
=& (\beta_2 + \beta_3 X_1) \Delta X_2
\end{aligned}$$

(c) If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$.

S&W Exercise 8.10 (iv)

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{i1} X_{2i} + u_i$$

Use Key Concept 8.1:

$$\Delta Y = f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2)$$

to show that

(a) $\Delta Y / \Delta X_1 = \beta_1 + \beta_3 X_2$ (effect of change in X_1, holding X_2 constant)

(b) $\Delta Y / \Delta X_2 = \beta_2 + \beta_3 X_1$ (effect of change in X_2, holding X_1 constant).

(c) If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$.

$$\begin{aligned}
\Delta Y =& f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2) \\
=& (\beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2(X_2 + \Delta X_2) + \beta_3(X_1 + \Delta X_1)(X_2 + \Delta X_2) + u) \\
& - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u) \\
=& \beta_1 \Delta X_1 + \beta_2 \Delta X_2 + \beta_3 \left(\Delta X_1 X_2 + X_1 \Delta X_2 + \Delta X_1 \Delta X_2\right) \\
=& (\beta_1 + \beta_3 X_2)\Delta X_1 + \left(\beta_2 + \beta_3 X_1\right) \Delta X_2 + \beta_3 \Delta X_1 \Delta X_2
\end{aligned}$$

Outline

1. S&W Exercise 8.2
2. S&W Exercise 8.10
3. SW Exercise 9.6
4. S&W Additional Empirical Exercise 8.1
5. S&W Additional Empirical Exercise 9.1
6. Cobb-Douglas Production Function

SW Exercise 9.6 (i)

Suppose that $n = 100$ i.i.d. observations for (Y_i, X_i) yield the following regression results:

$$\hat{Y} = \underset{(15.1)}{32.1} + \underset{(12.2)}{66.8}\ X,\ SER = 15.1, R^2 = 0.81$$

Another researcher is interested in the same regression, but he makes an error when he enters the data into his regression program: He enters each observation twice, so he has 200 observations (with observation 1 entered twice, observation 2 entered twice, and so forth).

(a) Using these 200 observations, what results will be produced by his regression program? (Hint: Write the "incorrect" values of the sample means, variances, and covariances of Y and X as functions of the "correct" values. Use these to determine the regression statistics.)

$$\hat{Y} = \underset{\_\_\_\_\_\_}{\_\_\_\_\_\_} + \underset{\_\_\_\_\_\_}{\_\_\_\_\_\_}\ X,\ SER = \_\_\_\_\_\_, R^2 = \_\_\_\_\_\_$$

(b) Which (if any) of the internal validity conditions are violated?

SW Exercise 9.6 (ii)

- Sample means do not change

$$\bar{X}' = \frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2n}\sum_{i=1}^{n} 2X_i = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\bar{Y}' = \frac{1}{2n}\sum_{i=1}^{2n} Y_i = \frac{1}{2n}\sum_{i=1}^{n} 2Y_i = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$$

- The OLS slope estimate does not change

$$\begin{aligned}\hat{\beta}_1{}' &= \frac{\sum_{i=1}^{2n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{2n}(X_i - \bar{X})^2} \\ &= \frac{\sum_{i=1}^{n} 2(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} 2(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \hat{\beta}_1\end{aligned}$$

- ... and therefore the OLS intercept will not change.

SW Exercise 9.6 (iii)

- The R^2 does not change.

$$R^{2\prime} = \frac{\sum_{i=1}^{2n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{2n}(Y_i - \bar{Y})^2} = \frac{\sum_{i=1}^{n} 2(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} 2(Y_i - \bar{Y})^2} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = R^2$$

- What about the SSR?

$$SSR' = \sum_{i=1}^{2n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n} 2(\hat{Y}_i - \bar{Y})^2 = 2SSR$$

- Therefore, the SER is as follows.

$$\hat{\sigma}_u' = \sqrt{\frac{1}{2n-2}SSR'} = \sqrt{\frac{1}{2n-2}2SSR} = \sqrt{\frac{n-2}{n-1}}\sqrt{\frac{1}{n-2}SSR} = \sqrt{\frac{n-2}{n-1}}\hat{\sigma}_u$$

If $n = 100$ and $\hat{\sigma}_u = 15.1$ then $\hat{\sigma}_u' = 15.02$

SW Exercise 9.6 (iv)

- Finally, what about the standard errors? Recall

$$\hat{\sigma}^2_{\beta_1} = \frac{1}{n}\frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2\hat{\sigma}^2_u}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}$$

Therefore,

$$\begin{aligned}
\hat{\sigma}^{2\prime}_{\beta_1} &= \frac{1}{2n}\frac{\frac{1}{2n-2}\sum_{i=1}^{2n}(X_i - \bar{X})^2\hat{\sigma}^{2\prime}_u}{\left(\frac{1}{2n}\sum_{i=1}^{2n}(X_i - \bar{X})^2\right)^2} \\
&= \frac{1}{2n}\frac{\frac{1}{2n-2}\sum_{i=1}^{n} 2(X_i - \bar{X})^2\left(\frac{n-2}{n-1}\right)\hat{\sigma}^2_u}{\left(\frac{1}{2n}\sum_{i=1}^{n} 2(X_i - \bar{X})^2\right)^2} \\
&= \left(\frac{1}{2}\frac{n-2}{n-1}\right)\frac{1}{n}\frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2\hat{\sigma}^2_u}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} = \left(\frac{n-2}{2n-2}\right)\hat{\sigma}^2_{\beta_1} \approx \frac{1}{2}\hat{\sigma}^2_{\beta_1} \\
\Rightarrow \hat{\sigma}'_{\beta_1} &= \sqrt{\frac{n-2}{2n-2}}\hat{\sigma}_{\beta_1} \approx \sqrt{\frac{1}{2}}\hat{\sigma}_{\beta_1} \text{ for large } n
\end{aligned}$$

If $n = 100$ and $\hat{\sigma}^2_{\beta_1} = 12.2$ then $\hat{\sigma}^{2\prime}_{\beta_1} = 8.58$

SW Exercise 9.6 (v)

- And

$$\hat{\sigma}^2_{\beta_0} = \frac{\hat{\sigma}^2_u \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Therefore,

$$\begin{aligned}
\hat{\sigma}^{2\prime}_{\beta_0} &= \frac{\hat{\sigma}^{2\prime}_u \sum_{i=1}^{2n} X_i^2}{2n \sum_{i=1}^{2n}(X_i - \bar{X})^2} \\
&= \frac{\frac{n-2}{n-1}\hat{\sigma}^2_u \sum_{i=1}^{n} 2X_i^2}{2n \sum_{i=1}^{n} 2(X_i - \bar{X})^2} \\
&= \frac{2}{4}\frac{n-2}{n-1}\frac{\hat{\sigma}^2_u \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{n-2}{2n-2}\hat{\sigma}^2_{\beta_0} \approx \frac{1}{2}\hat{\sigma}^2_{\beta_0} \\
\Rightarrow \hat{\sigma}^{\prime}_{\beta_0} &= \sqrt{\frac{n-2}{2n-2}}\hat{\sigma}_{\beta_0} \approx \sqrt{\frac{1}{2}}\hat{\sigma}_{\beta_0} \text{ for large } n
\end{aligned}$$

If $n = 100$ and $\hat{\sigma}^2_{\beta_1} = 15.1$ then $\hat{\sigma}^{2\prime}_{\beta_1} = 10.62$

SW Exercise 9.6 (vi)

Suppose that $n = 100$ i.i.d. observations for (Y_i, X_i) yield the following regression results:

$$\hat{Y} = \underset{(15.1)}{32.1} + \underset{(12.2)}{66.8}\, X,\ SER = 15.1, R^2 = 0.81$$

Another researcher is interested in the same regression, but he makes an error when he enters the data into his regression program: He enters each observation twice, so he has 200 observations (with observation 1 entered twice, observation 2 entered twice, and so forth).

(a) Using these 200 observations, what results will be produced by his regression program? (Hint: Write the "incorrect" values of the sample means, variances, and covariances of Y and X as functions of the "correct" values. Use these to determine the regression statistics.)

$$\hat{Y} = \underset{(10.62)}{32.1} + \underset{(8.58)}{66.8}\, X,\ SER = 15.02, R^2 = 0.81$$

(b) Which (if any) of the internal validity conditions are violated?

SW Exercise 9.6 (vii)

Suppose that $n = 100$ i.i.d. observations for (Y_i, X_i) yield the following regression results:

$$\hat{Y} = \underset{(15.1)}{32.1} + \underset{(12.2)}{66.8}\ X,\ SER = 15.1, R^2 = 0.81$$

Another researcher is interested in the same regression, but he makes an error when he enters the data into his regression program: He enters each observation twice, so he has 200 observations (with observation 1 entered twice, observation 2 entered twice, and so forth).

(a) Using these 200 observations, what results will be produced by his regression program? (Hint: Write the "incorrect" values of the sample means, variances, and covariances of Y and X as functions of the "correct" values. Use these to determine the regression statistics.)

(b) Which (if any) of the internal validity conditions are violated?
The data are no longer i.i.d. since half of the observations are precisely equal to the other half.

Outline

S&W Additional Empirical Exercise 8.1 (i)

Using the data set `TeachingRatings.dta` carry out the following exercises.

(a) Estimate a regression of *Course_Eval* on *Beauty*, *Intro*, *OneCredit*, *Female*, *Minority*, and *NNEnglish*.

(b) Add *Age* and *Age*^{2} to the regression.
- Is there evidence that *Age* has a nonlinear effect on *Course_Eval*?
- Is there evidence that *Age* has any effect on *Course_Eval*?

(c) Modify the regression in (a) so that the effect of *Beauty* on *Course_Eval* is different for men and women.
- Is the male–female difference in the effect of *Beauty* statistically significant?

(d) Professor Smith is a man. He has cosmetic surgery that increases his beauty index from one standard deviation below the average to one standard deviation above the average.
- What is his value of *Beauty* before the surgery?
- After the surgery?
- Using the regression in (c), construct a 95% confidence for the increase in his course evaluation.

(e) Repeat (d) for Professor Jones, who is a woman.

Outline

S&W Additional Empirical Exercise 9.1 (i)

A committee on improving undergraduate teaching at your college needs your help before reporting to the dean. The committee seeks your advice, as an econometric expert, about whether your college should take physical appearance into account when hiring teaching faculty. (This is legal as long as doing so is blind to race, religion, age, and gender.) You do not have time to collect your own data, so you must base your recommendations on the analysis of the data set `TeachingRatings.dta` described in Empirical Exercise AEE4.2 that has served as the basis for several of the additional empirical exercises. Based on your analysis of these data, what is your advice? Justify your advice based on a careful and complete assessment of the internal and external validity of the regressions that you carried out to answer the empirical exercises using these data in earlier chapters

S&W Additional Empirical Exercise 9.1 (ii)

- Threats to internal validity
  - Sources of bias/inconsistency in coefficients:
    1. Omitted variables
    2. Functional form misspecification
    3. Errors in variables (measurement error in the regressors)
    4. Sample selection
    5. Simultaneous causality
  - Incorrect standard errors:
    1. Heteroskedasticity
    2. Correlation in the error term across observations
- Threats to external validity
  - Can the results be generalized?
    1. Are the populations similar?
    2. Are the settings similar?
- Policy advice

Outline

Cobb-Douglas Production Function (i)

A "Cobb-Douglas" production function relates production (Q) to factors of production, capital (K), labor (L), and raw materials (M), and an error term u using the equation

$$Q = \lambda K^{\beta_1} L^{\beta_2} M^{\beta_3} e^u$$

where $\lambda, \beta_1, \beta_2$, and β_3 are production parameters. Suppose that you have data on production and the factors of production from a random sample of firms with the same Cobb-Douglas production function.

(a) How would you use regression analysis to estimate the production parameters?

Cobb-Douglas Production Function (ii)

A "Cobb-Douglas" production function relates production (Q) to factors of production, capital (K), labor (L), and raw materials (M), and an error term u using the equation

$$Q = \lambda K^{\beta_1} L^{\beta_2} M^{\beta_3} e^{u}$$

where $\lambda, \beta_1, \beta_2$, and β_3 are production parameters. Suppose that you have data on production and the factors of production from a random sample of firms with the same Cobb-Douglas production function.

(a) How would you use regression analysis to estimate the production parameters?

Note, by taking logs, we obtain

$$\ln(Q) = \ln(\lambda) + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + u$$

which is linear in the parameters, thus we can use OLS with this expression.

Cobb-Douglas Production Function (iii)

(b) Suppose that you would like to test that there are constant returns to scale in this industry. How would you do that?

(c) Is there a way to impose constant returns to scale in estimating the production parameters?

Cobb-Douglas Production Function (iv)

(b) Suppose that you would like to test that there are constant returns to scale in this industry. How would you do that?
$H_0 : \beta_1 + \beta_2 + \beta_3 = 1$, $H_A : \beta_1 + \beta_2 + \beta_3 \neq 1$,
Solution 1: F-test of the joint hypothesis.
Solution 2: Rewrite the equation to directly estimate the sum and then conduct a t-test on that coefficient.

$$
\begin{aligned}
\ln(Q) &= \ln(\lambda) + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + u \\
&= \ln(\lambda) + \beta_1 \ln(K) \pm \beta_2 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + u \\
&= \ln(\lambda) + (\beta_1 + \beta_2) \ln(K) + \beta_2 (\ln(L) - \ln(K)) + \beta_3 \ln(M) + u \\
&= \ln(\lambda) + (\beta_1 + \beta_2) \ln(K) \pm \beta_3 \ln(K) + \beta_2 (\ln(L) - \ln(K)) + \beta_3 \ln(M) + u \\
&= \ln(\lambda) + (\beta_1 + \beta_2 + \beta_3) \ln(K) + \beta_2 (\ln(L) - \ln(K)) \\
&\quad + \beta_3 (\ln(M) - \ln(K)) + u
\end{aligned}
$$

Thus a regression of $\ln(Q)$ on $\ln(K)$, $ln(L) - ln(K)$, and $ln(M) - ln(K)$, allows one to estimate the sum directly.

(c) Is there a way to impose constant returns to scale in estimating the production parameters?

Cobb-Douglas Production Function (v)

(b) Suppose that you would like to test that there are constant returns to scale in this industry. How would you do that?

(c) Is there a way to impose constant returns to scale in estimating the production parameters?
Yes, given the constraint $\beta_1 + \beta_2 + \beta_3 = 1$ we can make the following substitution: $\beta_3 = 1 - \beta_1 - \beta_2$

$$\begin{aligned}
\ln(Q) &= \ln(\lambda) + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + u \\
&= \ln(\lambda) + \beta_1 \ln(K) + \beta_2 \ln(L) + (1 - \beta_1 - \beta_2) \ln(M) + u \\
&= \ln(\lambda) + \beta_1 (\ln(K) - \ln(M)) + \beta_2 (\ln(L) - \ln(M)) \\
&\quad + \ln(M) + u \\
\implies \ln(Q) - \ln(M) &= \ln(\lambda) + \beta_1 (\ln(K) - \ln(M)) + \beta_2 (\ln(L) - \ln(M)) + u
\end{aligned}$$

Thus a regression of $\ln(Q/M)$ on $\ln(K/M)$, $In(L/M)$ allows one to estimate two factor shares under the constraint that all three sum to one.