Recitation 3

Sean Hyland

UN3412 Introduction to Econometrics

Spring 2020

Outline

1. Q1: S&W Exercise 4.1: Class Size and Test Scores
2. Q2: Natality and Education
3. Q3: S&W AEE 4.1: Earnings and Age
4. Q4: S&W AEE 4.2: Course Evaluations and Beauty
5. Q5: Stock Returns, continued

Outline

Q1: S&W Exercise 4.1: Class Size and Test Scores (i)

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS, \ R^2 = 0.08, \ SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (Hint: Review the formulas for the OLS estimators.)

(d) What is the sample standard deviation of test scores across the 100 class-rooms? (Hint: Review the formulas for the R^2 and SER.)

Q1: S&W Exercise 4.1: Class Size and Test Scores (ii)

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS, \;\; R^2 = 0.08, \;\; SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

$$E(TS|CS = 22) = 520.4 - 5.82 \times 22 = 392.36$$

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (Hint: Review the formulas for the OLS estimators.)

(d) What is the sample standard deviation of test scores across the 100 class-rooms? (Hint: Review the formulas for the R^2 and SER.)

Q1: S&W Exercise 4.1: Class Size and Test Scores (iii)

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS,\ R^2 = 0.08,\ SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

$$E(TS|CS = 23) - E(TS|CS = 19) = -5.82 \times (23 - 19) = -23.28$$

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (Hint: Review the formulas for the OLS estimators.)

(d) What is the sample standard deviation of test scores across the 100 class-rooms? (Hint: Review the formulas for the R^2 and SER.)

Q1: S&W Exercise 4.1: Class Size and Test Scores (iv)

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS,\ R^2 = 0.08,\ SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (Hint: Review the formulas for the OLS estimators.)

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \Rightarrow \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} = 520.4 + 5.82 \times 21.4 = 395.85$$

(d) What is the sample standard deviation of test scores across the 100 class-rooms? (Hint: Review the formulas for the R^2 and SER.)

Q1: S&W Exercise 4.1: Class Size and Test Scores (v)

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 520.4 - 5.82 \times CS,\ R^2 = 0.08,\ SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (Hint: Review the formulas for the OLS estimators.)

(d) What is the sample standard deviation of test scores across the 100 class-rooms? (Hint: Review the formulas for the R^2 and SER.)

$$s_y^2 = \frac{SST}{n-1} = \frac{\frac{SSR}{1-R^2}}{n-1} = \frac{(n-2)SER^2}{(1-R^2)(n-1)} = \frac{98 \times 11.5^2}{0.92 \times 99} = 142.30 \Rightarrow s_y = 11.5$$

Outline

1. Q1: S&W Exercise 4.1: Class Size and Test Scores
2. Q2: Natality and Education
3. Q3: S&W AEE 4.1: Earnings and Age
4. Q4: S&W AEE 4.2: Course Evaluations and Beauty
5. Q5: Stock Returns, continued

Q2: Natality and Education (i)

Let *KIDS* denote the number of children born to a woman, and let *EDUC* denote years of education for the woman. A simple model relating fertility to years of education is

$$KIDS = a + b \times EDUC + u,$$

where u is the unobserved residual.

(a) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

(b) Will simple regression of *KIDS* on *EDUC* uncover the ceteris paribus ('all else equal') effect of education on fertility? Explain.

Q2: Natality and Education (ii)

Let *KIDS* denote the number of children born to a woman, and let *EDUC* denote years of education for the woman. A simple model relating fertility to years of education is

$$KIDS = a + b \times EDUC + u,$$

where u is the unobserved residual.

(a) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

Income, age, and family background (such as number of siblings) are just a few possibilities.

Each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(b) Will simple regression of *KIDS* on *EDUC* uncover the ceteris paribus ('all else equal') effect of education on fertility? Explain.

Q2: Natality and Education (iii)

Let *KIDS* denote the number of children born to a woman, and let *EDUC* denote years of education for the woman. A simple model relating fertility to years of education is

$$KIDS = a + b \times EDUC + u,$$

where u is the unobserved residual.

(a) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

(b) Will simple regression of *KIDS* on *EDUC* uncover the ceteris paribus ('all else equal') effect of education on fertility? Explain.
Not if the factors we listed in part (i) are correlated with EDUC. Because we would like to hold these factors fixed, they are part of the error term. But if u is correlated with EDUC, then $E(u|EDUC) \neq 0$, and thus OLS Assumption (A2) fails. Recall

$$\hat{\beta}_1 \to_p \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X, \beta_0 + \beta_1 X + u)}{\mathrm{Var}(X)} = \beta_1 + \frac{\mathrm{Cov}(X, u)}{\mathrm{Var}(X)}$$

Outline

Q3: S&W AEE 4.1: Earnings and Age (i)

The data file `CPS12` contains data for full-time, full-year workers, age 25–34, with a high school diploma or B.A./B.S. as their highest degree. In this exercise, you will investigate the relationship between a worker's age and earnings. (Generally, older workers have more job experience, leading to higher productivity and earnings.)

(a) Run a regression of average hourly earnings (AHE) on age (Age). What is the estimated intercept? What is the estimated slope? Use the estimated regression to answer this question: How much do earnings increase as workers age by 1 year?

(b) Bob is a 26-year-old worker. Predict Bob's earnings using the estimated regression.
Alexis is a 30-year-old worker. Predict Alexis's earnings using the estimated regression.

(c) Does age account for a large fraction of the variance in earnings across individuals? Explain.

Outline

Q4: S&W AEE 4.2: Course Evaluations and Beauty (i)

The data file `TeachingRatings` contains data on course evaluations, course characteristics, and professor characteristics for 463 courses at the University of Texas at Austin. In this exercise, you will investigate how course evaluations are related to the professor's beauty.

(a) Construct a scatterplot of average course evaluations ($Course\_Eval$) on the professor's beauty ($Beauty$). Does there appear to be a relationship between the variables?

(b) Run a regression of average course evaluations on the professor's beauty. What is the estimated intercept? What is the estimated slope? Explain why the estimated intercept is equal to the sample mean of $Course\_Eval$. (Hint: what is the sample mean of Beauty?)

(c) Professor Watson has an average value of $Beauty$, while Professor Stock's value of $Beauty$ is one standard deviation above the average. Predict Professor Stock's and Professor Watson's course evaluations.

(d) Comment on the size of the regression's slope. Is the estimated effect of $Beauty$ on $Course\_Eval$ large or small? Explain what you mean by "large" and "small."

(e) Does Beauty explain a large fraction of the variance in evaluations across courses? Explain.

Outline

1. Q1: S&W Exercise 4.1: Class Size and Test Scores
2. Q2: Natality and Education
3. Q3: S&W AEE 4.1: Earnings and Age
4. Q4: S&W AEE 4.2: Course Evaluations and Beauty
5. Q5: Stock Returns, continued

Q5: Stock Returns, continued (i)

Return to question 3, part j, in the graded portion of the homework above. The R^2 values are notably different for the up market and down market sections of the data. Using the twoway scatter plot format, plot the two sections of the data separately, including in your plots the fitted line. With or without the information from the graphs, can you comment on or speculate about the difference in the R^2 values?

Q5: Stock Returns, continued (ii)

Return to question 3, part j, in the graded portion of the homework above. The R^2 values are notably different for the up market and down market sections of the data. Using the twoway scatter plot format, plot the two sections of the data separately, including in your plots the fitted line. With or without the information from the graphs, can you comment on or speculate about the difference in the R^2 values?

Two views:

(i) Despite the similar SERs between the two regressions, the dispersion in the up market graph around the fitted line is noticeably higher. It appears in the up market graph that there is particularly more dispersion at the higher return levels.

(ii) We can express R^2 as a function of SER and another term, highlighting a source of the discrepancy.

$$R_2 = 1 - \frac{SSR}{SST} = 1 - \frac{(N-2)SER^2}{SST} = 1 - \frac{SER^2}{\frac{SST}{N-2}} \approx 1 - \frac{SER^2}{s_y^2}$$