Econometrics 3220 - Spring 2020

Problem Set 2

This assignment is due on **Thursday, the 5th of March**. A hard copy has to be handed in at the **beginning** of lecture.

Question 1

This exercise asks you to prove some theoretical results. It is important to show all the steps in your proofs.

Consider the following simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad i = 1, \ldots, n$$

assuming $E(u|X) = 0$ and (Y_i, X_i), $i = 1 \ldots n$ are iid.

(a) What is the criterion for obtaining ordinary least squares estimators for β_0 and β_1? Write it using mathematical notation.

(b) Using the objective function specified in (a), derive the first order conditions and solve them simultaneously to recover the following formulae for the OLS estimators:

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X} \text{ and } \widehat{\beta_1} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

with: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

(c) What does it mean, in words, for an estimator to be unbiased? Prove that $\widehat{\beta_1}$ is an unbiased estimator under the given assumptions. The following steps will guide you through the proof:

(i) Using $\widehat{\beta_1} = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)Y_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$, plug in Y_i and simplify to show that

$$\widehat{\beta_1} = \beta_0(0) + \beta_1(1) + \frac{\sum(X_i - \bar{X})u_i}{\sum(X_i - \bar{X})^2}$$

.

You might be wondering why we are not using the expression for $\widehat{\beta_1}$ derived in (b). In fact, both the expressions are identical as you can easily show the two numerators are equal to each other $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)Y_i = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$. For unbiasedness, it is more convenient to use the expression in (i) which is just one of the several equivalent forms

(ii) Condition on X (which allows you to treat X as non-random) and take expectation on both sides. Use the assumption $E(u_i|X_i) = 0$ to show that

$$E(\widehat{\beta_1}) = \beta_1$$

.

(d) We further assume now that $Var(u|X) = \sigma^2$, assumed to be known. This is the assumption of homoskedasticity. Under this assumption, it can be shown that,

$$Var(\widehat{\beta_1}|X) = \frac{\sigma^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$$

Discuss intuitively how the variance of the estimator $\widehat{\beta_1}$ depends on the error variance and the variation in X.

Question 2

Consider the following Cobb-Douglas production function $Y_i = AK_i^{\beta_1} L_i^{\beta_2} e^{u_i}$ (where Y is output, A is the level of technology, K is the capital stock, L is the labor force and u is the error term)

Using data on (Y_i, K_i, L_i) for $i = 1, ..., n$ firms (where n is a large number) explain how you would test for constant returns to scale. Write down the regression you would run (how would you interpret the coefficients in this regression?) and the null hypothesis for constant returns to scale. Form the t-statistic for your test and show how you would get the numerator and the denominator for this t statistic. When would you reject the null of constant returns to scale at the 5% significance level?

Question 3

Suppose a researcher, using wage data on 200 randomly selected male workers and 240 female workers, estimates the OLS regression (standard errors in parentheses below coefficients)

$$\widehat{Wage} = \underset{(0.16)}{10.73} + \underset{(0.60)}{1.78} Male, R^2 = 0.09, SER = 3.8$$

where *Wage* is measured in dollars per hour and *Male* is a binary variable that is equal to 1 if the person is a male and 0 if the person is a female. Define the wage-gender gap as the difference in mean earnings between men and women.

(a) What is the estimated gender gap?

(b) Do men earn significantly more than women? Compute the p-value for testing the null hypothesis that there is no gender gap against the one-sided alternative that men earn more (refer to the normal table on pg. 804 of your textbook).

(c) Construct a 95% confidence interval for the gender gap.

(d) In the sample, what is the mean wage of women? What is the mean wage of men?

(e) Another researcher uses these same data but regresses *Wages* on *Female*, a variable that is equal to 1 if the person is female and 0 if the person is male. What are the coefficients (intercept and slope) calculated from this regression? What is the R^2 value? How do you interpret it?

(f) Does it make sense to assume that $E(u|Male) = 0$ where u is the error term in the model,

$$Wage = \beta_0 + \beta_1 Male + u$$

Discuss the implications of this assumption being violated. Based on your answer in (b), can you claim gender discrimination against females if $E(u|Male) \neq 0$? Explain.

Question 4

Consider the following multiple regression model

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

You want to consider certain hypotheses involving more than one parameter, and you know that the regression error is homoskedastic. You decide to test the joint hypotheses using the homoskedasticity-only F-statistics. For each of the cases below specify a restricted model and indicate how you would compute the F-statistic to test for the validity of the restrictions at the 1% level.

1. $\beta_1 = -\beta_2; \beta_3 = 0$
2. $\beta_1 + \beta_2 + \beta_3 = 1$
3. $\beta_1 = \beta_2; \beta_3 = 0$