Applied Econometrics
Midterm exam

Name: Debarshi Dutta

You can use your notes, the presentations, and the text. You should not work with other people.
Please send your work to me: derrell@fordham.edu
The midterm is due Thursday, October 22.

1. The table below gives the estimated coefficients and their p-values for the following regression. The p-values are for the two-sided test of statistical significance. D is a dummy variable.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 D + u$$

	Estimated coefficient	P-value
X1	-0.52	0.145
X2	1.23	0.023
D	0.44	0.078

a) What is the estimated value of y when x1 = 5, x2 = 10, and D = 1? (Beta0 = 2)

$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 D$
= 2 + (-0.52*5) + (1.23*10) + (0.44*1)
= **12.14**

b) Interpret the estimated value of beta1. Is beta1 statistically significant at the 10% level?

β_1 = -0.52.
The p-value of β_1 is 0.145 > 0.10. So β_1 is not statistically significant at the 10% level.

c) Suppose we had thought that beta1 should be negative. Is beta1 significantly negative at the 10% level?

β_1 = -0.52
The coefficient is negative.
The p-value of β_1 is 0.145 > 0.10.
So both the conditions are not satisfied.
Therefore, β_1 is not significantly negative

d) Interpret the estimated value of beta2. Is beta2 statistically significant at the 5% level?

β_1 = 1.23
The p-value of β_2 is 0.023 < 0.05. So β_2 is statistically significant at the 5% level.

e) What is the intercept when D = 0? What is the intercept when D = 1?

When D = 0,
The intercept = β_0 + β_1 + β_2 = **2.71**
When D = 1,
The intercept = β_0 + β_1 + β_2 + β_3 = **3.15**

f) You are considering the addition of another variable, x4, to the regression. You believe that x4 has a positive effect on y. You think that x4 is positively correlated with x2, but is not correlated with x1 or D. In the above reported regression, which coefficient (or coefficients) might be biased by the omission of x4? What is the direction of the expected bias?

x4 is correlated to x2 and has a positive effect on y. So there will be a bias.

Omitting the variable x4 will bias β_2. There is no bias for the other coefficients because x4 is not correlated with x1 or D.

$\beta_2 > 0$, so the direction of the expected bias will be positive.

2. Estimated regressions

a) $\hat{y} = 0.15 + 18.22 ln\, x$

Interpret the slope coefficient in the above estimated regression equation.

A one unit change of y leads to a change of 18.22 percent change of x

b) $\hat{ln y} = 1.23 - 0.147 ln\, x$

Interpret the slope coefficient in the above regression equation.

A one percent change of y leads to a change of negative 0.147 percent change of x

c) $\hat{y} = 0.13 + 1.27x - 0.35xD$

In the above regression equation, D is a dummy variable.

What is the estimated slope coefficient when D = 0?

The slope coefficient is 1.27

What is the estimated slope coefficient when D = 1?

The slope coefficient is (1.27-0.35) = 0.92

d) $\hat{y} = 0.25 + 0.17x + 0.05x^2$

What is the estimated slope coefficient when x = 4?

When x = 4,

$\hat{y} = 0.25 + (0.17).4 + (0.05).4^2$

= 1.73 = constant

So, the slope coefficient is zero

3. A scatter plot of y and x shows a u-shaped relationship. Write a regression equation that will fit this relationship.

The U-shaped curve denotes a polynomial. So, the equation can be, $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$, which is a parabola

4. You believe x1 and x2 both have positive effects on y. You also believe that the effect of x1 on y grows stronger as x2 increases. Write a regression equation to fit these relationships.

$y = \beta_0 + \beta_1 x_1^n + \beta_2 x_2 + u$ (n > 1)

5. OLS regression assumes that the mean of the error terms is zero. What must you do to assure that this assumption is met?

The average of the error terms being zero ensures that the model is unbiased. The relationship in the OLS model has to be linear for the mean of the error terms to be zero. Random chance should determine the error terms.

6. Consider the following regression equation. Write the null and alternative hypotheses for the test of the overall significance of the regression. Why would you perform this test?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

H_0: There is no relationship between the independent variables (x_1, x_2 and x_3) and the dependent variable (y)
H_1: There is a relationship between the dependent variable and the independent variables.

Testing the null hypothesis (H_0) leads to either rejecting H_0 or failure to reject it, which consequently establishes whether a relationship does or doesn't exist between the set of independent variables and the dependent variable. When H_0 is rejected it means that there is a relationship between the set of independent variables and the dependent variable, that is, the alternative hypothesis H_1 is accepted. Failure to reject H_0 signifies there is in fact no relationship between the independent and dependent variables. This results in verifying statistical assumptions.

7. Consider the linear probability model from your last homework assignment. Use the same data. Add the first lag of the dependent variable:

$$Recess_t = \beta_0 + \beta_1 Spread_{t-1} + \beta_2 Baa_{t-1} + \beta_3 Recess_{t-1} + u$$

Report the estimated regression equation. Comment on the statistical significance of the slope coefficients. Is this model better than the model without $Recess_{t-1}$? Why or why not?

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.47763	0.11855	4.03	0.0002
SPREAD1	1	-0.26815	0.05273	-5.09	<.0001
BAA1	1	0.07485	0.07019	1.07	0.2905
RECESS1	1	0.20675	0.10192	2.03	0.0470

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.51886	0.11974	4.33	<.0001
SPREAD1	1	-0.30199	0.05129	-5.89	<.0001
BAA1	1	0.11187	0.06949	1.61	0.1126

RECESS = -0.269SPREAD1 + 0.0749BAA1 + 0.207RECESS1 + 0.478

Working at the 5% level, the slope coefficient of *SPREAD1* that is, the lag of *SPREAD* is < 0.0001. So it is statistically significant. RECESS1 is also statistically significant (p-value = 0.0470)

If we compare both models the only difference is the inclusion of $RECESS_{t-1}$ as a dependent variable. Its inclusion does not change the statistical significance of the other variable coefficients. $RECESS_{t-1}$ is statistically significant. So the second model with $RECESS_{t-1}$ included will be better.

8. What is the topic of your research paper? What is the dependent variable? What are the independent variables? Where will you get the data? This question will not be graded but please provide answers.

Topic- Factors behind the cause of obesity

Hypothesis- Counties with a high share of the population receiving food assistance (SNAP, free lunch for students program) as well as low access to healthy, affordable food and fitness facilities will have higher rates of obesity.

Data Source - https://www.ers.usda.gov/webdocs/DataFiles/48731/DataDownload.xls?v=9288.7