# Midterm Assessment

EC 204: Empirical Economics 2

Due on: Thursday July 30, 2020 by 10:00AM (EST)

**Instructions**

The objective of this assessment is to test your knowledge of multiple regression estimation and inference as well as your ability to code in STATA. You have 48 hours to complete this exam. You may refer to the course material when completing this assessment. **You cannot discuss this assessment with any individual(s). Doing so will be considered cheating.**

You are responsible for making sure that you understand each of the four (4) questions clearly. If not, be sure to ask me any questions that might arise. If you use any STATA commands/options/statistical methods that have not been discussed in class, briefly describe what they are and justify why you chose to use them.

All relevant data files are provided in the "Midterm" folder (in .dta format as always). In addition, there is a pdf with some relevant statistical tables. **Be sure to submit your .do file along with your responses to the exam questions. These must all be compiled in one file.**

Good luck!

## Question 1

Suppose you estimate the following model by OLS:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

$wage$ : hourly age in dollars
$educ$ : years of education
$exper$ : years of experience

You obtain the following fitted model using STATA, where standard errors are given in parenthesis

$$\widehat{wage} = 3.5 + 0.9 educ + 1.5 exper$$
$$(2.0) \quad (0.7) \qquad (0.5)$$
Number obs. : 523
$$R^2 = 0.45$$

For the following questions, make use to the relevant statistical tables in the "Midterm" folder. If you find that the degrees of freedom exceed 120, then use the critical values in the $\infty$ row.

(a) At the 1% level, conduct a test of the hypothesis that $\beta_2 = 0$ against the alternative that $\beta_2 > 0$. Formally write out the null hypothesis. Show how you calculate the appropriate test statistic, and make careful note of the degrees of freedom, the critical value of your test statistic, and whether or not you reject the null.

(b) Now, at the 5% level conduct a test of the null that experience and education *jointly* have no effect on wage against the alternative that this is not true. Formally state the null and alternative hypotheses. Show how you compute the appropriate test statistic, and make note of degrees of freedom, critical values, and whether or not you reject the null.

## Question 2

The effects of campaign expenditures on voting outcomes are of interest to both economists and political scientists. The dataset VOTE.DTA contains data for 173 two-party races for the U.S. House of Representatives in 1988. The variable *voteD* contains the percentage of votes received by the Democratic candidate. Expenditures by the two candidates (in thousands of dollars) are given by *expendR* and *expendD*. The variable *prtystrD* measures the Democratic party's general strength in the district as a percentage of the 1988 Presidential vote in that district that went to the Democratic candidate.

(a) Estimate the relationship between voting outcomes and expenditures using an appropriate econometric model. Provide an argument supporting the model and estimator that you use (as well as standard errors). How effective are campaign expenditures in explaining voting (interpret your results)?

(b) Is campaign spending an important determinant of vote shares? Explain.

(c) Does an equal increase in spending in a campaign (by both candidates) favor Democrats, favor Republicans, or have no effect? How strong is this result statistically? Does this conclusion affect your choice of econometric model?

(d) Is an additional $\$1,000$ of campaign spending by Democrats on Congressional races more effective at increasing the Democratic share of the vote in districts that tend to vote Democratic (based on Presidential voting)?

## Question 3

Use the data in HTV.DTA to answer this question. The dataset includes information on wages, education, parents' education, and several other variables for 1,230 working men in 1991.

(a) What is the range of the `educ` variable in the sample? What percentage of men completed 12th grade but no higher grade? Do the men or their parents have, on average, high levels of education?

(b) Suppose the only determinants of education are mother and father education. To that end, estimate the following two models

$$\text{educ} = \widetilde{\beta}_0 + \widetilde{\beta}_1 \text{motheduc} + v \tag{1}$$
$$\text{educ} = \beta_0 + \beta_1 \text{motheduc} + \beta_2 \text{fatheduc} + u \tag{2}$$

Report the results for regression (2). How much sample variation in `educ` is explained by parents' education? Interpret the coefficient on `fatheduc`.

(c) Now add the variable `abil` (a measure of cognitive ability) to regression (2) and report the results. Does "ability" help explain variations in education, even after controlling for parents' education? Explain.

(d) Now add the square of ability, $\texttt{abil}^2$, to the regression in part (c). Characterize the partial effect of ability on education, controlling for parent education. At what level of "ability" does the partial effect of `abil` become zero? (Note that `abil` can take on negative values.)

## Question 4

Use data in VOLAT.DTA for this exercise. The variable $rsp500$ is the monthly return on the Standard & Poor's 500 stock market index, at an annual rate. (This includes price changes as well as dividends.) The variable $i3$ is the return on the three-month T-bills, and $pcip$ is the percentage change in industrial production; these are also at an annual rate.

(a) Consider the equation

$$rsp500_t = \beta_0 + \beta_1 pcip_t + \beta_2 i3_t + u_t$$

what signs do you think $\beta_1$ and $\beta_2$ should have? Why?

(b) Estimate the equation in $(a)$ by OLS and report the results. Interpret the signs and magnitudes of the coefficients.

(c) Which of the variables is statistically significant?

(d) Does your finding from $(c)$ imply that the return on the S&P500 is predictable? Explain.