Problem Set 3

EC 204: Empirical Economics 2

Due on: Thursday August 6, 2020 by 10:00am (EST)

True/False/Uncertain (20 points)

Instructions: For the each of the following statements, indicate whether it is true, false, or if the assertion is indeterminate. In all cases, defend your answer with a concise explanation. **To earn full credit, you must include an explanation with your answer.**

1. Since x^2 is a function of x, we will be faced with perfect multicollinearity if we attempt to use both x and x^2 as regressors.
2. Measurement error in the dependent variable is more serious than measurement error in the independent variables.
3. A sufficient condition for a "good" instrument is that it satisfies the relevance condition.
4. If a variable follows a random walk with drift, the best forecast of the variable tomorrow is the value of the variable today.

Longer Question (10 Points)

1. (**Categorical Variables**) Suppose a sample of adults is classified into groups 1, 2, and 3 on the basis of whether their education stopped at the end of elementary school, high school, or university, respectively. The relationship is

$$y = \beta_1 + \beta_2 D_2 + \beta_3 D_3 + \varepsilon$$

where y is income, $D_i = 1$ for those in group i and zero for all others.

   (a) Explain why D_1 is not included in the regression.

   (b) In terms of the parameters of the model, what is the expected income of people whose education stopped at the end of university? What is the expected income of people whose education stopped at the end of elementary school?

   (c) Suppose some respondents who only finished elementary school were embarrassed about their lack of education and lied, claiming that they graduated from high school. What would be the impact of this lie on the estimates of β_2 and β_3?

Computational Exercises (20 Points)

1. (**Wooldridge, Chapter 8, C13)** Use the data in FERTIL2.DTA to answer this question

   (a) Estimate the model

   $$children = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 educ + \beta_4 electric + \beta_5 urban + u$$

   and report the usual and heteroskedasticity-robust standard errors. Are the robust standard errors always bigger than the nonrobust ones?

   (b) Add the three religious dummy variables and test whether they are jointly significant. What are the p-values for the nonrobust and robust tests?

   (c) From the regression in (b), obtain the fitted values $\widehat{y}$ and the residuals $\widehat{u}$. Regress $\widehat{u}^2$ on $\widehat{y}$ and $\widehat{y}^2$ and test the joint significance of the two regressors. Verify that heteroskedasticity is present in the equation for *children*.

   (d) Would you say the heteroskedasticity you found in (c) is practically important?

2. Use the data in GPA2.DTA for this exercise

   (a) Consider the equation

   $$colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u$$

   where *colgpa* is cumulative college grade point average; *hsize* is size of high school graduating class, in hundreds; *hsperc* is academic percentile in graduating class; *sat* is combined SAT score; *female* is a binary gender variable; and *athlete* is a binary variable, which is one for student athletes. What are your expectations for the coefficients in this equation? Which ones are you unsure about?

   (b) Estimate the equation in (a) and report the results. What is the estimated GPA differential between athletes and non-athletes? Is it statistically significant?

   (c) Drop *sat* from the model and reestimate the equation. Now, what is the estimated effect of being an athlete? Discuss why the estimate is different than that obtained in (b).

   (d) In the model from part (a), allow the effect of being an athlete to differ by gender and test the null hypothesis that there is no ceteris paribus difference between women athletes and women non-athletes.