EC421 take-home exam

Submission

You must submit your answers to this exam **as a Word document** (or in a similar/comparable format) **on Canvas before 11:59pm (PST)** on Wednesday, **18 March 2020**.

If you submit this exam, its grade will replace your lowest homework grade. If you do not submit this exam, your grade will remain as it currently is on Canvas.

Materials

You can use our class materials (notes, assignments, book), online materials, and anything else.

You cannot discuss this exam with other people.

Cheating

Again: You cannot discuss this exam with any other person or work with any other person. If we suspect you of cheating, we will give you a **zero for the entire course** and may report you to the dean.

If two exams are suspected of cheating, the individuals of both exams will receive zeros.

Points

There are 108 total points on the exam.

Section 1: "True" or "False and explain"

13 questions, 4 points per question (52 points)

Instructions: Answer whether the statement is **true** or **false**. If the statement is **false**, you must **explain why** the statement is incorrect. Your explanation should be 1–2 sentences. Incomplete or too-long answers will both lose points. Finally, there are only two true statements.

`00.` **Yes or No:** Do you promise not to cheat on this exam? We will only grade exams that answer "Yes" to this question. (This is not a true/false question, and there are no points for this question.)

`01. T/F/Explain` If our model has a lagged explanatory variable and u_t has first-degree autocorrelation, then OLS is biased and inconsistent for estimating the coefficients, *e.g.*, $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + u_t$.

`02. T/F/Explain` *Omitted-variable bias* means coefficient estimates will be smaller than the true value.

`03. T/F/Explain` *Mean stationarity* implies that the mean of a variable cannot differ from zero.

`04. T/F/Explain` Exogeneity means that our disturbances' variances are constant.

`05. T/F/Explain` The standard error of an estimator can safely be ignored.

`06. T/F/Explain` In the Rubin causal model, $y_{1,i}$ refers to individual i's outcome when she is in the treatment group.

`07. T/F/Explain` When the *p*-value is greater than 0.05, we conclude that the null hypothesis is true.

`08. T/F/Explain` OLS is never consistent for a regression model with a lagged dependent variable.

`10–12` In the regression equation

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Female}_i + \beta_3 \text{Education}_i \times \text{Female}_i + u_i$$

> `09. T/F/Explain` The model allows that the wage returns to education to differ for women and men.
>
> `10. T/F/Explain` If *Ability* is correlated with *Education* and affects *Wage*, then omitting *Ability* will cause our estimate of β_1 to be higher than the true value of β_1.
>
> `11. T/F/Explain` If *Height* (and individual's height) is correlated with *Gender* and does not affect *Wage*, then omitting *Height* will bias our coefficient estimate for β_2.

12. `T/F/Explain` In the presence of measurement error in our explanatory variables (as defined in class), our coefficient estimates are biased downward.

13. `T/F/Explain` The first stage in two-stage least squares (2SLS) and insturmental variables (IV) is when we regress the outcome variable on the instrumental variable.

Section 2: Short-answer questions

14 questions, 4 points per question (56 points)

Instructions: Provide a short but thorough answer/explanation to each question.

14. For the following regression model, we do not allow the effect of education to depend on an individual's level of experience.

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Experience}_i + u_i$$

Write out (as an equation) a model that allows the effect of an individual's experience to vary by her level of education. (Equations have variables, subscripts, math, and Greek letters.)

15. What does it mean for a relationship between two variables to be "spurious"?

16–19 Suppose that we want to know how sleep affects test scores, *e.g.*,

$$\text{Score}_i = \beta_0 + \beta_1 \text{Sleep}_i + u_i$$

> 16. Define/explain both of the requirements for a valid *instrumental variable* in this setting.
>
> 17. Explain why we may want to use instrumental variables (or two-stage least squares) in this setting. (Hint: What problem are IV and 2SLS trying to solve?)
>
> 18. Give an example of an omitted variable in this situation.
>
> 19. Now defined Job_i as a variable that describes whether person i has an off-campus job. Explain whether Job_i is a valid instrument for Sleep_i (explain whether it satisfies each requirement for a valid instrument).

20. Suppose we have a regression model with lagged dependent variable, *e.g.*,

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + u_t$$

What must be true about the relationship between u_t and u_{t-1} for OLS to consistently estimate β_2?

21–24 Suppose we want to know how health depends on gender and employment status. We define the variables as:

- `health` is a score between 0 and 100 for an individual's health.
- `female` is a binary variable `0` or `1` for whether the individual is female.
- `employed` is a binary variable `0` or `1` for whether the individual is employed.

Consider the output from the following regression.

```
#>                  Estimate Std. Error
#> (Intercept)       24.6427     0.2844
#> female            20.7793     0.4053
#> employed          30.7402     0.3989
#> female:employed    1.0206     0.5687
```

21. Write out the model (as an equation) that we've estimated.

    Note: We do not want numbers in equation—we're looking for something with our dependent variable as a function of the explanatory variables, βs, *etc*.

22. Interpret the coefficient on `female:employed`.

23. Interpret the coefficient on `female`.

24. Is the interaction statistically significant? Explain.

25–27 Suppose that we have two groups of people. In the first group ("Group 1"), everyone exercises every day. In the second group ("Group 2"), no one ever exercises. We want to know the effect of exercise on an individual's weight.

25. Explain why comparing the average weight in Group 1 to the average weight in Group 2 will probably not yield an unbiased/consistent estimate of the effect of exercise on weight.

26. Define the fundamental problem of causal inference and explain how it relates to the current example.

27. Propose a method/technique (that we've discussed in class) that would produce a consistent estimate of the effect of exercise on weight. Explain why this technique/method is the right answer..