EC 421
Midterm

13 February 2020

Full Name ←

UO ID ←

No phones, calculators, or outside materials.

A. True or false

40 points

Note: You do not need to explain to your answers **in this section**. There will be no partial credit.

01. [T/F] (**2pts**) In the presence of measurement error (as defined in class), our coefficients are always biased toward zero.

02. [T/F] (**2pts**) *Omitted-variable bias* results in your coefficient estimates being smaller than the true value (for example, $\hat{\beta}_1 < \beta_1$, on average).

03. [T/F] (**2pts**) Heteroskedasticity does not bias OLS's estimates of the coefficients.

04. [T/F] (**2pts**) If our disturbances have different variances, then we have a violation of exogeneity.

05. [T/F] (**2pts**) The asymptotic properties of an estimator (for example: consistency) have to do with an estimator's behavior as the sample size approaches infinity.

06. [T/F] (**2pts**) In the regression equation

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

we allow the wage effect of an individual's *Experience* to vary by her level of *Education*.

07. [T/F] (**2pts**) For a Goldfeld-Quandt test, if SSE_1 and SSE_2 are equal, we will generally reject the null hypothesis and conclude there is significant evidence of heteroskedasticity.

08. **[T/F]** (**2pts**) Specifying the wrong functional form for your regression model can lead to heteroskedasticity.

09. **[T/F]** (**2pts**) Our assumption of exogeneity requires that $E[u_i \mid x_i] \neq 0$.

10. **[T/F]** (**2pts**) Omitted-variable bias only affects OLS's *unbiasedness* and not its consistency.

11.–13. In the regression equation

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

11. **[T/F]** (**2pts**) The model assumes that the wage returns to education are the same for women and men.

12. **[T/F]** (**2pts**) If *Ability* is correlated with *Education* and affects *Wage*, then omitting *Ability* will bias our estimate of β_1.

13. **[T/F]** (**2pts**) 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.

14. **[T/F]** (**2pts**) Our assumption of exogeneity is critical for OLS's unbiasedness.

15. **[T/F]** (**2pts**) If an estimator is unbiased, then it it is consistent.

16. **[T/F]** (**2pts**) For random variables X and Y: $\text{plim}(X \times Y) = \text{plim}(X) \times \text{plim}(Y)$

17. **[T/F]** (**2pts**) In the presence of heteroskedasticity, WLS (weighted least squares) is an unbiased estimator of the coefficients (the β_j).

18. **[T/F]** (**2pts**) In the presence of heteroskedasticity, WLS (weighted least squares) is less efficient than OLS for estimating the coefficients (the β_j).

19. **[T/F]** (**2pts**) Whereas e_i is an unobservable population parameter, u_i is observable.

20. **[T/F]** (**2pts**) The main problem with omitted-variable bias is that it biases our standard errors, which causes our inference to be wrong.

Short answer

60 points

Note: You will typically need to explain/justify your answers in this section.

21. (**3pts**) Imagine we are testing the null hypothesis H_0 $\beta_1 = 3$ against the alternative hypothesis H_a $\beta_1 \neq 3$. If the p-value is 0.9, what should we conclude?

22. (**3pts**) Define the concept of the median.

23. (**3pts**) Define the term "variance."

24. (**3pts**) What does $E[X]$ tell us about the random variable X?

25. (**3pts**) For the model

$$\log(Q_i) = 12 - 0.3\log(P_i) + u_i$$

interpret the slope. *Note:* P denotes price (in dollars), and Q refers to quantity (in "units").

26. (**4pts**) Define the term "standard error."

27. (**3pts**) Explain why we care about the standard error of an estimator—for example, the standard error of the OLS estimator $\hat{\beta}_1$.

28. (**3pts**) For the model

$$\log(Q_i) = 12 - 0.3P_i + u_i$$

interpret the slope. *Note:* P denotes price (in dollars), and Q refers to quantity (in "units").

29. (**4pts**) What are the **two** requirements for omitted variable bias?

30. (**3pts**) Why do we care about heteroskedasticity?

31. Suppose we run the regression

$$\text{Health}_i = \beta_0 + \beta_1 \text{Income}_i + u_i$$

but the true model is actually

$$\text{Health}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Stress}_i + u_i$$

Where $\beta_2 < 0$ (*i.e.*, stress is bad for health).

Also: Recall that the probability limit of our OLS-based estimate for β_1 is

$$\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(\text{Income}_i, \text{Stress}_i)}{\text{Var}(\text{Income}_i)}$$

a. (**2pts**) If income and stress are positively correlated, will our regression be biased? If so, will it overestimate or underestimate the true effect of income? Briefly explain your answer.

b. (**2pts**) If income and stress are negatively correlated, will our regression be biased? If so, will it overestimate or underestimate the true effect of income? Briefly explain your answer.

c. (**2pts**) If income and stress are uncorrelated, will our regression be biased? If so, will it overestimate or underestimate the true effect of income? Briefly explain your answer.

32. In the regression equation

$$\text{Score}_i = \beta_0 + \beta_1 \text{GPA}_i + \beta_2 \text{Class}_i + \beta_3 \text{GPA}_i \times \text{Class}_i + u_i$$

let Score_i denote individual i's test score, GPA refers to i's GPA, and Class_i describes whether i attends class (0 for 'no', or 1 for 'yes').

a. (**2pts**) Interpret the coefficient β_0. Explain why this coefficient is a bit strange to interpret.

b. (**2pts**) Interpret the coefficient β_1.

c. (**2pts**) Interpret the coefficient β_3.

d. (**2pts**) Suggest an omitted variable that could cause β_1 to be biased. Explain.

32. (continued)

e. (**3pts**) Imagine we are concerned about heteroskedasticity. Walk me through the steps for running a White test for heteroskedasticity (regressions that we would run, the null hypothesis, alternative hypothesis, *etc.*).

f. (**2pts**) Suppose our White test has a p-value of 0.041. What is our conclusion? Explain.

33. (**3pts**) Draw a plot where the disturbances are **homoskedastic**.
Your plot should have u on the y-axis and x on the x-axis.

34. (**3pts**) Draw a plot where the disturbances are **heteroskedastic**.
Your plot should have u on the y-axis and x on the x-axis.

35. (**3pts**) Draw a plot where the disturbances are **not exogenous**.
Your plot should have u on the y-axis and x on the x-axis.

Extra credit

6 points

EC$_1$ [T/F] (**2pts**) Omitted-variable bias has nothing to do with whether we interpret regression estimates as causal.

EC$_2$ (**2pts**) Write down the regression equation that we would estimate in the following line of `R` code (*i.e.*, the equation with βs).

```
lm(crime ~ police + income + police:income, data = city_df)
```

EC$_3$ (**2pts**) Draw a plot of **heteroskedastic disturbances** for which the Breusch-Pagan test would fail to find significant evidence of heteroskedasticity.