EC 421
Midterm

13 February 2020

Full Name ← SOLUTIONS

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?

We fail to reject Ho. (We do not have sufficient evidence to reject $\beta_1 = 3$.)

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

The median is the middle observation — equal numbers of observations on each side.

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

Variance is the amount a random variable deviates from its mean.

Also: $Var(X) = E[(X - \mu_x)^2]$

$\mu_x = E(X)$

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

Its mean.

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").

The slope tells us that a 100% (1%) increase in price will reduce quantity by 30% (0.3%)

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

The standard error is the standard deviation of an estimator's distribution.

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

SE tells us how precise (or uncertain) our estimate is.

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").

For a one-dollar increase in price, we anticipate a 30% decrease in quantity.

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

1. The omitted variable correlates with an included regressor (an "X").
2. The omitted variable affects our outcome (y).

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

Heteroskedasticity biases our standard errors — messing up inference.
(Also: reduces efficiency for estimating coefficients.)

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

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

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.

Our regression estimate will be biased / consistent.
We will underestimate the true effect of income on health:

$$\text{plim}\ \hat{\beta}_1 = \beta_1 + \underbrace{(-)\frac{(+)}{(+)}}_{(-)} < \beta_1$$

so our estimate will tend to be too small.

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.

Our regression estimate will be biased / consistent.
We will overestimate the true effect of income on health:

$$\text{plim}\ \hat{\beta}_1 = \beta_1 + \underbrace{(-)\frac{(-)}{(+)}}_{(+)} > \beta_1$$

so our estimate will tend to be too large

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.

Our regression will be unbiased / consistent.

$$\text{plim}\ \hat{\beta}_1 = \beta_1 + \underbrace{(-)\frac{0}{(+)}}_{=0} = \beta_1$$

(consistent)

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.

β_0 tells us the avg. score for individuals who do not attend class and have a GPA of 0.

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

β_1 tells us the effect of GPA on test scores for individuals who do not attend class.

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

β_3 tells us the difference of the effect of GPA on scores between the class-goers and those who do not attend class.

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

Ability: It affects test scores and correlates with GPA — the requirements for omitted variable bias.

32. (continued)

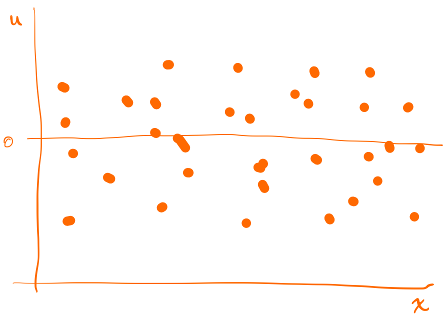
e. (**3pts**) Imagine we are concerned about heteroskedasticity. Walk me through the steps for running a White test for heteroskedasticity (regressions that we would run, the null hypothesis, alternative hypothesis, *etc.*).

1. Run regression in the equation.
2. Using residuals from step 1: run regression of squared residuals on all terms in original regression PLUS their squares and their interactions.
3. Use R^2 from step 2 to test H_0: Homoskedasticity vs. H_A: Heteroskedasticity
   $\alpha_1 = \alpha_2 = \cdots \alpha_p = 0$ $\quad$ $\alpha_i \neq 0$ for some i

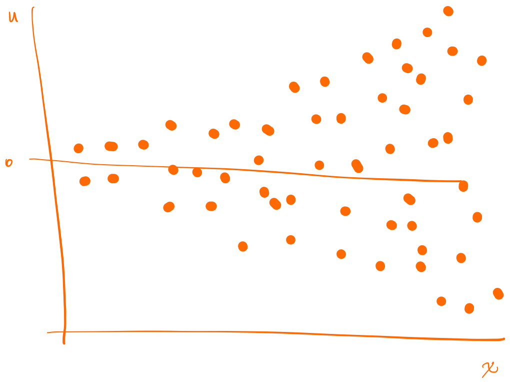
f. (**2pts**) Suppose our White test has a p-value of 0.041. What is our conclusion? Explain.

Reject H_0 and conclude we have statistically significant evidence of heteroskedasticity at the 5% level.

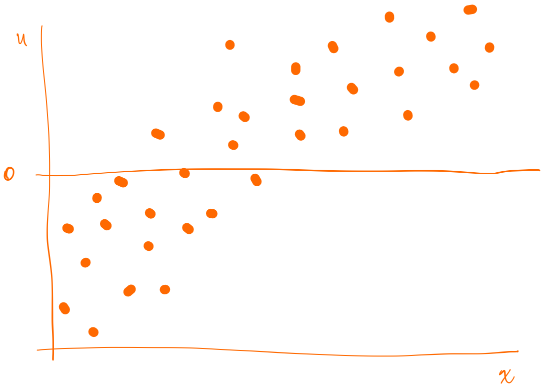
33. (**3pts**) Draw a plot where the disturbances are **homoskedastic**.
Your plot should have u on the y-axis and x on the x-axis.



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.



Basically: $E[u|x] \neq 0$ for some x

Extra credit

6 points

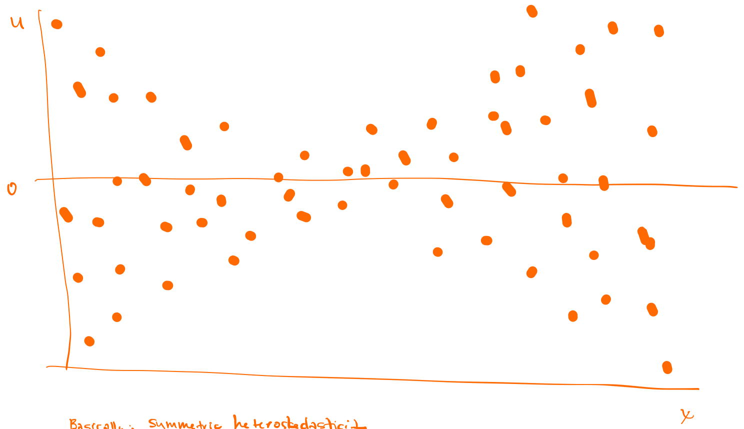
EC$_1$ **[T/F]** (**2pts**) Omitted-variable bias has nothing to do with whether we interpret regression estimates as causal.

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)
```

$$\text{Crime}_i = \beta_0 + \beta_1 \text{Police}_i + \beta_2 \text{Income}_i + \beta_3 \text{Police}_i \times \text{Income}_i + u_i$$

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



Basically: Symmetric heteroskedasticity.