Name: _____

In-class midterm, EC421

120 points possible

Part 1: True or false (50 points)

Note: In this section, select the correct answer. You do not need to explain your answer.

1. **[T/F]** (2.5pts) Exogeneity essentially says that the disturbance must be independent of your explanatory variables.
2. **[T/F]** (2.5pts) Exogeneity also implies that the disturbance must be independent of the outcome variable.
3. **[T/F]** (2.5pts) If an estimator is biased, then it is also inconsistent.
4. **[T/F]** (2.5pts) A variable will cause omitted-variable bias in OLS estimates when the the following things are true:
    1. The is omitted from the regression.
    2. The omitted variable correlates with one of the included regressors.
5. **[T/F]** (2.5pts) Adding additional variables always increases R^2.
6. **[T/F]** (2.5pts) Homoskedasticity means that $\text{Var}(u_i) = \text{Var}(u_j)$ for all individuals i and j in the population.
7. **[T/F]** (2.5pts) If the sum of the squared residuals increases (holding the sample size constant), then the standard errors will also increase.
8. **[T/F]** (2.5pts) If your model of interest is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

then the regression for the corresponding White test for heteroskedasticity should estimate

$$e_i^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2 + v_i$$

9. **[T/F]** (2.5pts) Suppose you conduct a t test and find that the p-value is 0.065.

   True or false: You would conclude that the coefficient is significantly different from zero at the five-percent level.

10. **[T/F]** (2.5pts) Consider the following model of how income depends upon gender (an indicator for whether i is male) and ethnicity (an indicator for whether individual i Hispanic).

$$\text{Income}_i = \beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Hispanic}_i + \beta_3 \text{Male}_i \times \text{Hispanic}_i + u_i$$

    True or false: The expected income for a non-Hispanic male is $\beta_0 + \beta_3$.

11. **[T/F]** (2.5pts) Omitted-variable bias also affects OLS's consistency.

12. **[T/F]** (2.5pts) If you use heteroskedasticity-robust standard errors, and the disturbance is actually homoskedastic, then your standard errors are biased.

13. **[T/F]** (2.5pts) Consider the following econometric model of the number of car accidents.

$$\text{Accidents}_i = \beta_0 + \beta_1 \text{Traffic}_i + \beta_2 \text{Precipitation}_i + u_i$$

    True or false: In the model, effect of precipitation on the number of car accidents depends on the amount of traffic.

14. **[T/F]** (2.5pts) Measurement error (as defined in class) biases coefficients downward.

15. **[T/F]** (2.5pts) In the 'real world', you should assume the disturbance is homoskedastic.

16. **[T/F]** (2.5pts) In the model

$$\text{Employed}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Age}_i + u_i$$

    assume that Employed_i is a binary indicator for whether individual i is employed.

    True or false: β_1 is the change in the *probability* that an individual is employed that results from a 1-unit increase in education (all else equal).

17. **[T/F]** (2.5pts) Suppose you observe that whenever the variable X increases, the variable Y also increases.

    True or false: Either (1) X causes Y or (2) Y causes X.

18. **[T/F]** (2.5pts) Consistency tells us about the behavior of an estimator as the sample size approaches infinity.

19. **[T/F]** (2.5pts) In the model

$$\log \text{Income}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Age}_i + u_i$$

If $\hat{\beta}_1 = 0.57$, then, on average, a one-percent increase in education generates a 0.57-percent increase in income (all else equal).

20. **[T/F]** (2.5pts) Measurement error in our outcome variable causes attenuation bias.

Questions continue on the next page.

Part 2: A single model (20 points)

Note: The questions in this section all use the following model, where we regress the length of individuals' commute (in minutes) on a binary variable for whether individual i is a college graduate.

$$(\text{Commute length})_i = \beta_0 + \beta_1 (\text{College Graduate})_i + u_i$$

Suppose that the estimates are $\hat{\beta}_0 = 26$ and $\hat{\beta}_1 = -14$.

Use these estimates (and the definition of the model) to answer the following four questions. Your answers for questions 21–23 should be a single number.

21. (5pts) What is the average commute length for individuals who are not college graduates?

$\hat{\beta}_0 = 26$

22. (5pts) What is the average difference in commute length between college graduate and non-graduates (college graduates minus non-graduates)?

$\hat{\beta}_1 = -14$

23. (5pts) What is the average commute length for individuals who graduated college?

$\hat{\beta}_0 + \hat{\beta}_1 = 26 - 14 = 12$

24. (5pts) Provide an example of a variable that would likely cause omitted variable bias for our estimate of β_1 above. Explain how the variable likely meets the requirements to cause omitted-variable bias.

One example: Income.

Requirements:

1. Omitted variable affects the outcome variable. (Income might affect commute length.)
2. Omitted variable correlates with included variable. (Income correlates with college grad.)

Part 3: Short-answer questions (40 points)

Note: In this section briefly answer the questions/prompts in 1–3 short (and complete) sentences. We will deduct points for excessively long answers.

25. (5pts) Weighted least squares (WLS) essentially divides observations by the standard deviation of their disturbance (*e.g.*, dividing by σ_i). Explain the intuition for how this can increase efficiency (relative to OLS).

    Individuals with larger variances in their disturbances are "less understood" by the model—in other words, the model has more uncertainty about these individuals. Consequently, if we focus more on (upweight) individuals with low-variance disturbances, we can learn more, as these are, by definition, the individuals that the model understands.

26. (5pts) Explain (1) what measurement error is and (2) the problems it causes for OLS regression.

    As defined in class, measurement error occurs when one of our variables is mis-measured. In other words, we do not observe the *actual* variable of interest but rather the variable *plus* some *noise*.

    When an explanatory variable has measurement error, OLS is biased toward zero (attenuation bias).

    Note: We focused on measurement error in the explanatory variable.

27. (5pts) What problems are caused when individuals' disturbances have different variances?

    When individuals' disturbances have different variances—heteroskedasticity—OLS's standard errors are biased (which affects inference like hypothesis tests and confidence intervals), and OLS is less efficient (no longer the most efficient linear, unbiased estimator).

28. (5pts) Which of the linear-regression assumptions does omitted-variable bias violate? Briefly explain your answer.

Omitted-variable bias violates exogeneity.

When we omit a variable that affects our outcome, that variable shows up in the disturbance.

If that omitted variable correlates with an included regressor, then the disturbance is correlated with a regressor.

This correlation between the disturbance and a regressor violates exogeneity.

29. (5pts) Why do we typically prefer the White test to the Goldfeld-Quandt test?

The White test is able to detect more types of heteroskedasticity and makes fewer assumptions.

30. (5pts) Provide an example of when individuals disturbances would be correlated. Explain.

Individuals disturbances might correlate when they live in the same household. They will share many similar "shocks" (weather, local economy, employment, kids) that are in the disturbance. Thus, their disturbances will likely correlate.

31. (5pts) Which issue causes bigger issues in econometrics: omitted-variable bias or measurement error? Explain your answer.

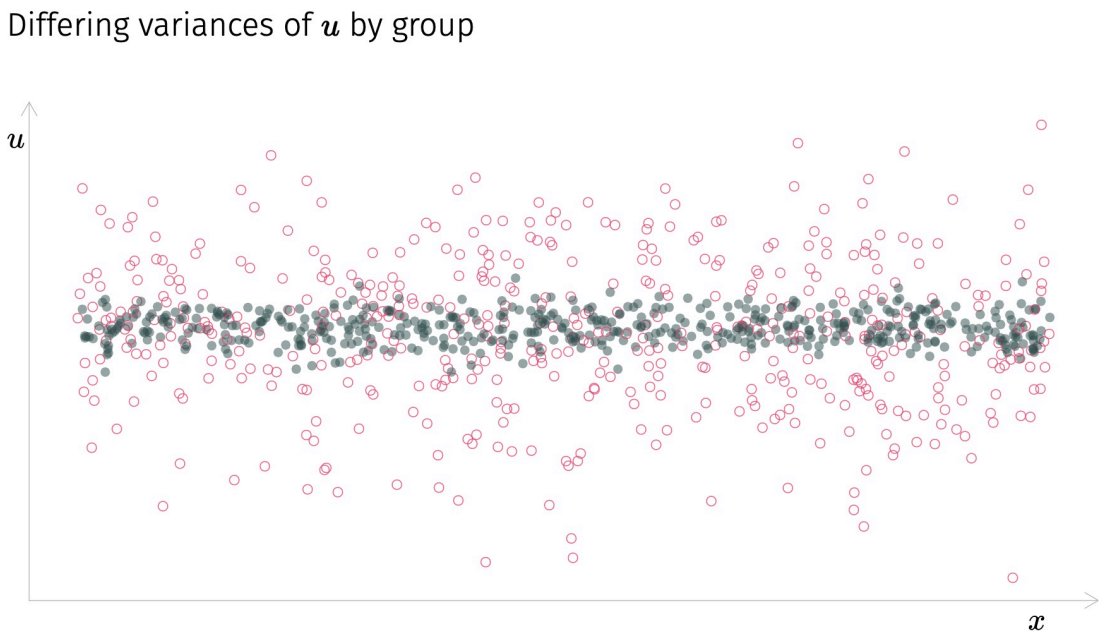
Both issues are BIG. Omitted-variable bias can cause you to get the magnitude and even sign of the estimated parameter wrong. Measurement error can bias you to zero. Ignoring either one is likely to cause important bias.

I'm just looking for good reasoning.

32. (5pts) Draw a single figure that illustrates heteroskedastic disturbances such that

- There are two groups (you can draw x's and o's).
- The disturbances *within a group* are homoskedastic.
- The two groups' disturbances have different variances.

Example from the notes...



Part 3: Long-ish answer (10 points)

Note: Answer the prompt with 1–3 paragraphs.

33. (10 pts) Imagine you want to estimate the effect of policing on the number of crimes.

    - **Step 1** (3 points) Explain how you could use OLS regression to estimate this effect.
    - **Step 2** (3 points) Write out the econometric model that you would estimate to capture this effect. (It should be a mathematical model with βs—not R code.)
    - **Step 3** (4 points) Explain what 'real-world' concerns you have for obtaining an unbiased or consistent estimate for the true effect of policing on crime.

Looking for good reasoning with complete answers.