Name: Key

In-class midterm, EC421

120 points possible

1 True or false (50 points; 20 questions)

Note In this section, select the correct answer (true or false). You do not need to explain your answer.

F 1. (2.5 points) **[T/F]** Ordinary least squares (OLS) regression is biased for estimating the coefficients of a linear model when the disturbances are heteroskedastic.

T 2. (2.5 points) **[T/F]** One of the "standard" assumptions for OLS regression is that the disturbances are uncorrelated with each other.

F 3. (2.5 points) **[T/F]** If our disturbances have different variances, then we have a violation of exogeneity.

F 4. (2.5 points) **[T/F]** In the regression model

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

the effect of education on income depends on the individual's age.

F 5. (2.5 points) **[T/F]** The White test for heteroskedasticity for the above model would run the following regression:

$$e_i^2 = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Age}_i^2 + \beta_3 \text{Education}_i \times \text{Age} + v_i.$$

F 6. (2.5 points) **[T/F]** While OLS is biased when we violate exogeneity, it is still consistent.

T 7. (2.5 points) **[T/F]** The square of a variable is equivalent to interacting the variable with itself.

F 8. (2.5 points) **[T/F]** In the following model, if $\hat{\beta}_1 = 0.5$, then a one-unit increase in X_i is associated with a 0.5-unit increase in Y_i:

$$\log(y)_i = \beta_0 + \beta_1 x_i + u_i.$$

F 9. (2.5 points) **[T/F]** Weighted least squares (WLS) upweights individuals with higher variance disturbances and downweights individuals with lower variance disturbances.

F 10. (2.5 points) **[T/F]** If the disturbance correlates with an explanatory variable, then we have a violation of the homoskedasticity assumption.

T 11. (2.5 points) **[T/F]** An estimator can be consistent without being unbiased.

T 12. (2.5 points) **[T/F]** Adding additional variables to a regression model will always increase the R^2.

T 13. (2.5 points) **[T/F]** Disturbances are unobservable, while residuals are observable.

F 14. (2.5 points) **[T/F]** Measurement error in the explanatory variable tends to bias the OLS estimator downward.

T 15. (2.5 points) **[T/F]** For the Goldfeld-Quandt test, the null hypothesis is that the variances of the disturbances are equal across groups.

F 16. (2.5 points) **[T/F]** Consistency tells us about the behavior of an estimator when we take an infinite number of samples with the same sample size.

F 17. (2.5 points) **[T/F]** In the following regression model, the expected income for a non-female student is $\beta_0 + \beta_1$

$$\text{Income}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Student} + \beta_3 \text{Female}_i \times \text{Student}_i + u_i.$$

F 18. (2.5 points) **[T/F]** A p-value of 0.5 means that we can reject the null hypothesis that the coefficient is equal to zero at the 5% level.

T 19. (2.5 points) **[T/F]** One problem of the Goldfeld-Quandt test for heteroskedasticity is that it fails to detect certain patterns of heteroskedasticity.

F 20. (2.5 points) **[T/F]** Omitted-variable bias occurs when we omit a variable from a regression.

2 Multiple choice (20 points; 5 questions)

Note In this section, check (✓ or ×) **all** correct answers. You do not need to explain your answer.

21. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following statements are part of the "standard" assumptions for OLS regression?

    ☑ $E[u_i|X_i] = 0$ ☐ $\text{Var}(u_i) = 0$ ☑ $\text{Var}(X_i) > 0$ ☑ $\text{Cov}(u_i, u_j) = 0$

22. (4 points) **[Multiple choice]** Choose *all* correct answers:

    In the presence of heteroskedasticity, which of the following statements are true?

    ☐ OLS is unbiased for the coefficients.

    ☐ WLS is biased for the coefficients.

    ☑ WLS is the best linear unbiased estimator.

    ☑ OLS is biased for the standard errors.

23. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following scenarios necessarily causes OLS to be biased for the coefficients?

    ☐ heteroskedasticity ☐ correlated disturbances ☑ violating exogeneity ☑ measurement error

24. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following models *violates* OLS's requirement of linearity?

    ☐ $\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u$

    ☐ $y = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u}$

    ☑ $y = \beta_0 + \beta_1 x_1^{\beta_2} + \beta_3 x_2 + u$

    ☑ $y = \beta_0 + \beta_1 x_1 \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + u$

25. (4 points) **[Multiple choice]** Choose *all* correct answers:

    What can we "do" in the presence of heteroskedasticity?

    ☑ Check the specification. ☐ Look for omitted variables.

    ☑ Use het.-robust standard errors. ☑ Estimate the model using WLS.

3 Short answer (50 points; 10 questions)

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

26. (5 points) Explain how OLS regression defines the "best-fit" line.

The line that minimizes the SSE = $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y - [\hat{\beta}_0 + \hat{\beta}_1 x_i]\right)^2$

27. (5 points) Define the concept of a *standard error*.

The standard deviation of an estimator's distribution.

28. (5 points) Explain how the phrase *correlation is not causation* relates to the concept of *omitted-variable bias*.

Omitted-variable bias is one reason why correlation does not imply causation.
Independent things may correlate due to omitted variables.

29. (5 points) Write down a simple linear regression model (actual variables; not just y and x), and provide an example of a variable that could cause omitted variable bias. Explain how the variable satisfies the requirements for omitted-variable bias.

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

Omitted variable: Pollution_i

Pollution exposure varies w/. income and affects health.

30. (5 points) In class we showed that for included regressor x_1 and excluded regressor x_2, the coefficient on x_1 has probability limit

$$\text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)} \times \beta_2$$

Using this formula, explain which direction you would expect the OLS estimator to be biased when we regress *income* on *education* and omit *ability*.

Equation was incorrect:

All answers were considered "correct."

31. (5 points) Compare and contrast consistency and unbiasedness.

- Both cons. and unbias. are ways to think about an estimator's behavior — particularly whether the estimator tends to give the "right" answer (the parameter of interest).
- Consistency asks whether the estimator's distribution collapses to a "spike" at the desired param, as sample size $\to \infty$.
- Unbiasedness asks whether the estimator on average (exp. value) gives the right answer for a fixed and finite sample size.

32. (5 points) For the regression model below, suppose we estimate $\hat{\beta}_0 = 9.4$ and $\hat{\beta}_1 = -0.5$. Interpret the slope coefficient in the context of the model.

$$\log(\text{Quantity}_i) = \beta_0 + \beta_1 \log(\text{Price}_i) + u_i$$

All else equal, we expect quantity to decrease 50% when price increases 100%.

33. (5 points) For the regression model below, suppose we estimate $\hat{\beta}_0 = 55.4$, $\hat{\beta}_1 = 0.3$, $\hat{\beta}_2 = 12.1$, and $\hat{\beta}_3 = 0.1$. Interpret the coefficient on the interaction term.

$$(\text{Years Lived})_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Female}_i + \beta_3 \text{Income}_i \times \text{Female}_i + u_i$$

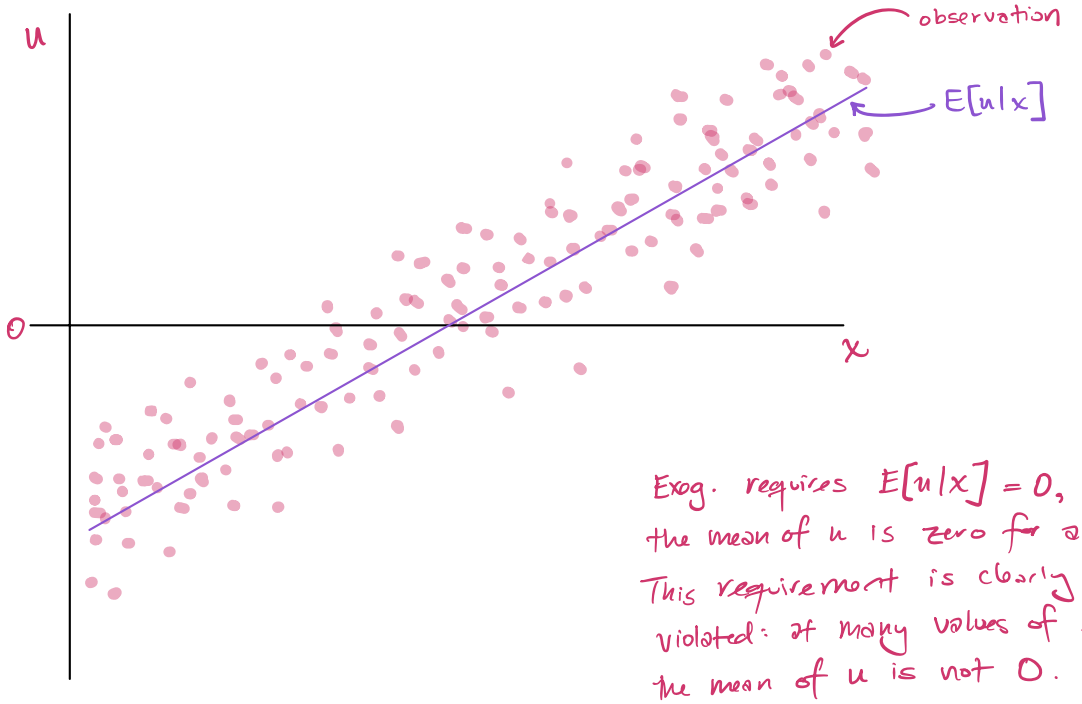
Note: Income is measured in thousands of dollars; *Female* is a binary indicator variable.

For each $1,000 increase income, life expectancy increases 0.1 years more for females than non-females (holding all else equal).

34. (5 points) Imagine you are running the White test for heteroskedasticity. You notice that the coefficient on x_1^2 is statistically significant. What does this tell you about the presence of heteroskedasticity? Explain.

A stat. significant coefficient on x_1^2 suggests the disturbance is likely heteroskedastic.

35. (5 points) Draw an example of a scatterplot that would violate the assumption of exogeneity. Briefly explain why the scatterplot violates the assumption.



Exog. requires $E[u|x] = 0$, ie, the mean of u is zero for all x. This requirement is clearly violated: at many values of x, the mean of u is not 0.