EC 421

Midterm

12 February 2019

Full Name ← KEY

UO ID ←

No phones, calculators, or outside materials.

A. True/False, Multiple Choice, and Fill in the Blank

40 points

Note: You do not need to explain to your answers **in this section**.

01. **[T/F]** (**2pts**) For the model $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$, we interpret β_1 as the percentage change in y due to a one-unit increase in x.

02. **[T/F]** (**2pts**) For the model $\log(y_i) = \beta_0 + \beta_1 \log(x_i) + u_i$, we interpret β_1 as the percentage change in y due to a one-unit increase in x.

03. **[T/F]** (**2pts**) The model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ violates are linearity assumption by allowing y and x to have a *nonlinear* relationship.

04. **[T/F]** (**2pts**) When we use heteroskedasticity-robust standard errors, we still use OLS to estimate the coefficients (the β's).

05. **[T/F]** (**2pts**) Omitted-variable bias occurs whenever we omit a variable from a regression.

06. **[T/F]** (**2pts**) Heteroskedasticity is a violation of our exogeneity assumption.

07. **[T/F]** (**2pts**) In the presence of heteroskedasticity, OLS provides unbiased standard errors.

08. **[T/F]** (**2pts**) Measurement error biases the OLS estimates for the coefficients toward zero.

09. **[T/F]** (**2pts**) Omitted-variable bias causes OLS estimates for the coefficients to be biased, but OLS is a consistent estimator for the coefficients.

10. **[T/F]** (**2pts**) OLS is biased for models with lagged explanatory variables.

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

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

13. [T/F] (2pts) For random variables X and Y: $E[X \times Y] = E[X] \times E[Y]$.

14. **[Multiple choice]** (**2pts**) In the presence of heteroskedasticity, which of the following is true?

A. OLS and WLS are biased, but OLS is less biased.
B. OLS is biased. WLS is unbiased.
C. OLS is unbiased. WLS is biased.
D. OLS and WLS are unbiased, but OLS is more efficient.
E. OLS and WLS are unbiased, but WLS is more efficient.

15. **[Multiple choice]** (**2pts**) Which of the following can lead to heteroskedasticity?

A. Adding a lagged outcome variable
B. Misspecification
C. Measurement error
D. Disturbances with different variances

16. **[Fill in the blanks]** (**4pts**) The **expected value** of an estimator is the mean of the estimator's distribution (for a set sample size n), whereas the estimator's **probability limit** describes how the estimator behaves as n approaches infinity.

17. **[Fill in the blank]** (**2pts**) If the estimator $\hat{\beta}_1$ is unbiased for β_1, then $E\left[\hat{\beta}_1\right] = \beta_1$.

18. **[Fill in the blank]** (**4pts**) If our significance level is 0.05, and our p-value is 0.07, then we Fail to reject the null hypothesis.

B. Short Answer

60 points

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

19. **(3pts)** What is the difference between u_i and e_i?

u_i is the population disturbance, i.e., $u_i = y_i - (\beta_0 + \beta_1 x_i)$

e_i is the regression residual, i.e., $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$
(which depends upon our sample)

20. **(3pts)** **Briefly explain** what we mean by "the standard error of an estimator"?

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

21. One of our assumptions is that the errors are homoskedastic.

i. **(2pts)** Formally write down this assumption. *In other words:* Write out this assumption using expected values and/or variances of random variables (*i.e., use 'math'*).

$\text{Var}(u_i | X) = \sigma^2$ for all i (will accept $\text{Var}(u_i) = \sigma^2$)

Also acceptable: $\mathbb{E}[u_i^2 | X] = \sigma^2$ or $\mathbb{E}[u_i^2] = \sigma^2$

ii. **(3pts)** Informally (in words): What does this assumption mean?

This assumption means that there is no relationship between the variance of our disturbances and our explanatory variables.

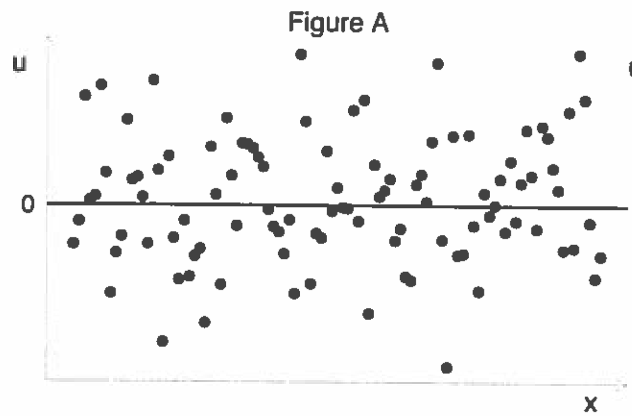
Alternative: We have constant variance in our disturbances.

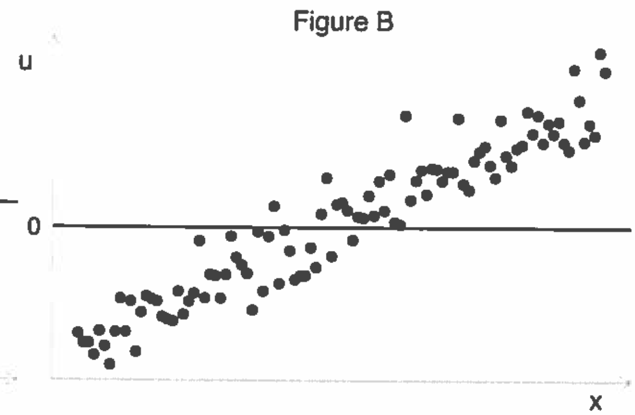
iii. **(2pts)** What does rejecting H_0 in a Goldfeld-Quandt test tell us about this assumption?

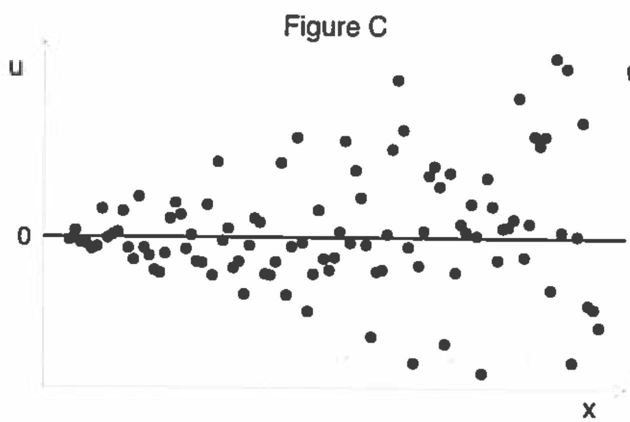
Rejecting Ho for a G-Q test implies statistically significant evidence of heteroskedasticity — violating this assumption.

Alternative: Our two groups differ in the variance of their disturbances, which violates this assumption.

22. For each of the four figures below (depicting x_i and u_i), answer whether the figure suggests a violation of homoskedasticity. **Briefly explain** each of your answers.







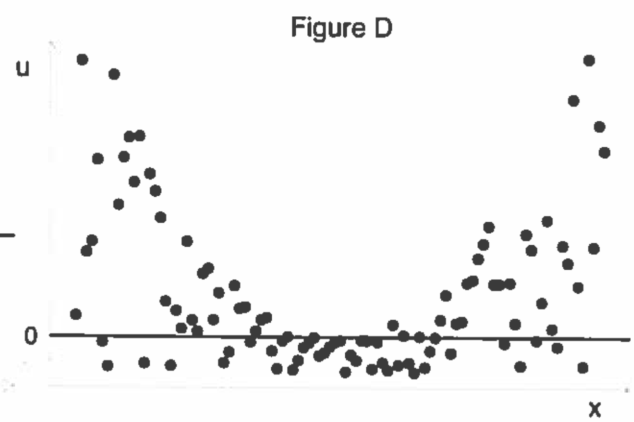


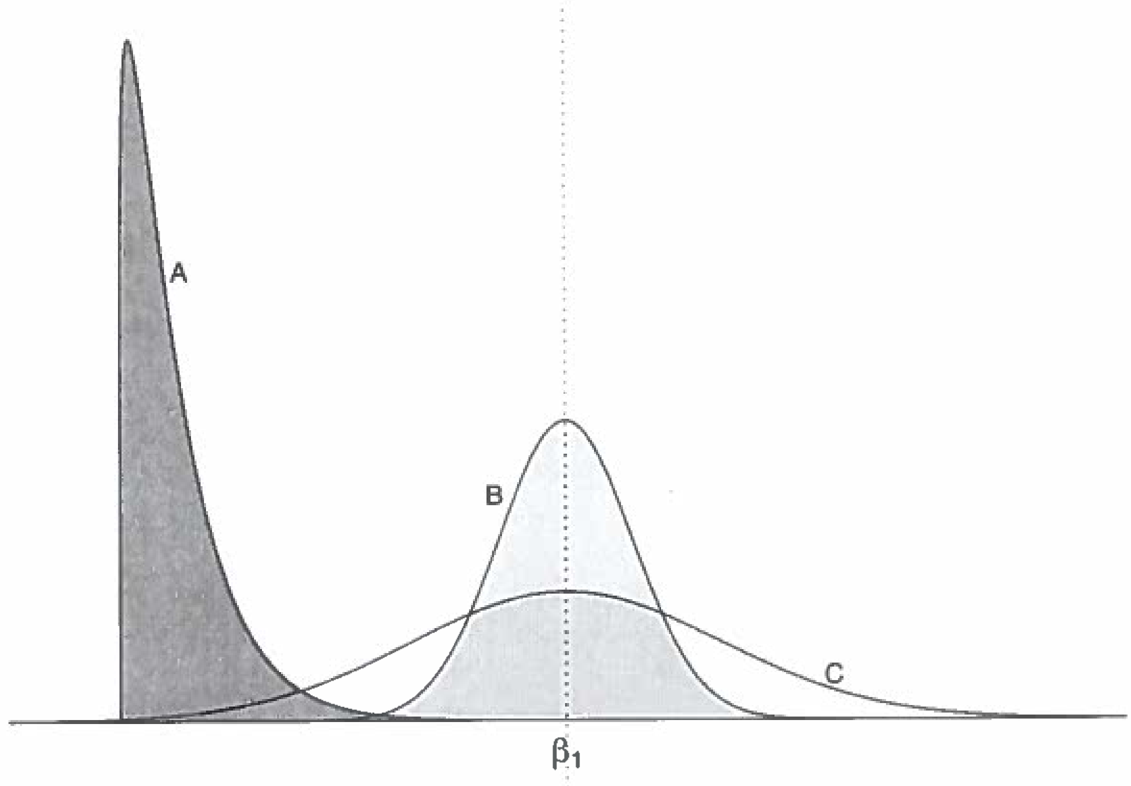
Figure A: (1pt) No violation — disturbances appear to have constant variance.

Figure B: (1pt) No violation: variance is constant.

Figure C: (1pt) Violation: Variance increases w/. X.

Figure D: (1pt) Violation: variance decreases then increases w/. X.

23. The figure below shows the distributions of three estimators for β_1: A, B, and C.



i. (**4pts**) Which estimators, if any, are unbiased? ***Briefly* explain** your answer.

B & C: their means are at the parameter they estimate (β_1).

ii. (**2pts**) Which estimator is the most efficient? ***Briefly* explain** your answer.

A: tightest around its mean. (lowest variance)

iii. (**2pts**) Does this figure tell us anything about the consistency of estimator **A**? **Explain.**

No. Consistency is about how the estimator behaves as $n \to \infty$.
We know nothing about $n \to \infty$ from this picture.

24. We want test for heteroskedasticity using a Breusch-Pagan test. We start by estimating the model

$$\text{Crime}_i = \beta_0 + \beta_1 \text{Police}_i + \beta_2 \text{Wages}_i + u_i$$

i. (4pts) Walk me through the necessary steps for completing the Breusch-Pagan test. Do not estimate anything—just outline the procedure, step by step.

1. Estimate the model above.
2. Regress squared residuals (from (1)) on expl. variables.
3. Calculate LM test stat. using R^2 from (2).
4. Compare LM test stat to critical value or obtain p-value (using χ^2_2) — not necessary for full points.
5. Reject or fail to reject Ho: Homoskedasticity.

ii. (3pts) Suppose we calculate a Breusch-Pagan test statistic of $\text{LM} = 12.3$, which has a p-value of approximately 0.031. Complete the Breusch-Pagan test (with conclusions).

Reject Ho: Homosk.

Conclude: We have stat. significant evidence at 5% level that our disturbances are heteroskedastic.

iii. (2pts) How would a White test differ from the Breusch-Pagan test?

We would include squared terms and interactions ~~[illegible]~~ in our regression of e_i^2.

25. Consider the model

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

where $\textbf{Experience}_i$ measures individual i's experience in the workforce (in years), and $\textbf{Female}_i$ is a binary variable for whether individual i is female.

i. (**4pts**) What is the interpretation of β_1?

For a one-year increase in experience, we expect wages to increase by β_1 dollars, holding all else constant.

ii. (**3pts**) What is the interpretation of β_2?

β_2 gives the difference in wages between women and men, holding experience fixed.

Alternatively: β_2 gives the "effect" of being female on wages, holding experience constant.

iii. (**2pts**) Suppose you think the effect of *experience* changes depending upon whether the individual is female. Write down (**1**) a model that could test this hypothesis and (**2**) the null and alternative hypotheses that you would test.

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

$H_0: \beta_3 = 0 \quad H_A: \beta_3 \neq 0$

26. (**2pts**) Give an example of a variable that would likely produce measurement error. Briefly explain your answer.

(impossible?)
Happiness: Difficult to actually measure; we tend to use proxies.

GDP: Hard to measure; noisily measured.

27. Consider the time-series model

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_3 \text{Income}_{t-2} + u_t$$

i. (**2pts**) What are the interpretations of β_1 and β_2 and how do they differ?

β_1: the effect of additional income today on births, today.

β_2: the effect of additional income yesterday on births today.

Differ: one is contemporaneous effect (β_1); other is lagged effect (β_2).

ii. (**2pts**) What is the *total* effect of income on births?

$\beta_1 + \beta_2 + \beta_3$

28. Suppose that we take a sample. Our observations in this sample come from two groups: **Group A** has very little noise/variation in its disturbances; **Group B** has *a lot* of noise/variation in its disturbances.

i. (**2pts**) **[T/F]** OLS ignores these differences in the variances of the disturbances.

ii. (**2pts**) **Explain the intuition** for how WLS takes advantage of differences in these disturbances' variances.

WLS upweights low-variance observations (obs. w/. low σ_i^2) and downweights "noisy" observations (high-variance disturbances).

29. We have in mind the model

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

but we cannot observe *Ability*. Instead, we run the regression

$$\text{Income}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{Education}_i + e_i$$

i. (3pts) What two conditions are required for omitted-variable bias to bias our estimate of the effect of education on income?

1. Ability affects income.
2. Ability is correlated w/. education.

ii. (2pts) If we assume that ability positively affects income, and ability is positively correlated with education, then will our estimate of β_1 be biased upward or downward?

Upward. $\left(\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(\text{educ}, \text{ability})}{\text{Var}(\text{educ})} \right)$

not necessary

iii. (2pts) How does your answer to (ii.) change if ability positively affects income but is negatively correlated with education?

Then OLS would be downward biased for β_1.

C. Extra Credit

10 points

EC$_1$ **[T/F]** **(2pts)** If you fail to reject the null hypothesis in a Goldfeld-Quandt test, then you can conclude your disturbances are homoskedastic.

EC$_2$ For the R code and output below,

```
# The regression
ec_reg <- lm(problems ~ money, data = extra_credit)
# The results
tidy(ec_reg)
```

```
#> # A tibble: 2 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)   21.7     0.187        116. 1.12e-106
#> 2 money          3.00    0.00335      896. 1.45e-193
```

i. (2pts) Write down the model that we are estimating.

$$\text{problems}_i = \beta_0 + \beta_1 \text{money}_i + u_i$$

ii. (2pts) Interpret the output for the intercept—including its statistical significance.

The intercept tells us that people w/out money ($\text{money}_i = 0$) have, on average, 21.7 problems.

EC3 For each of the four figures below (depicting x_i and u_i), answer whether the figure suggests a violation of exogeneity. **Briefly explain** each of your answers.

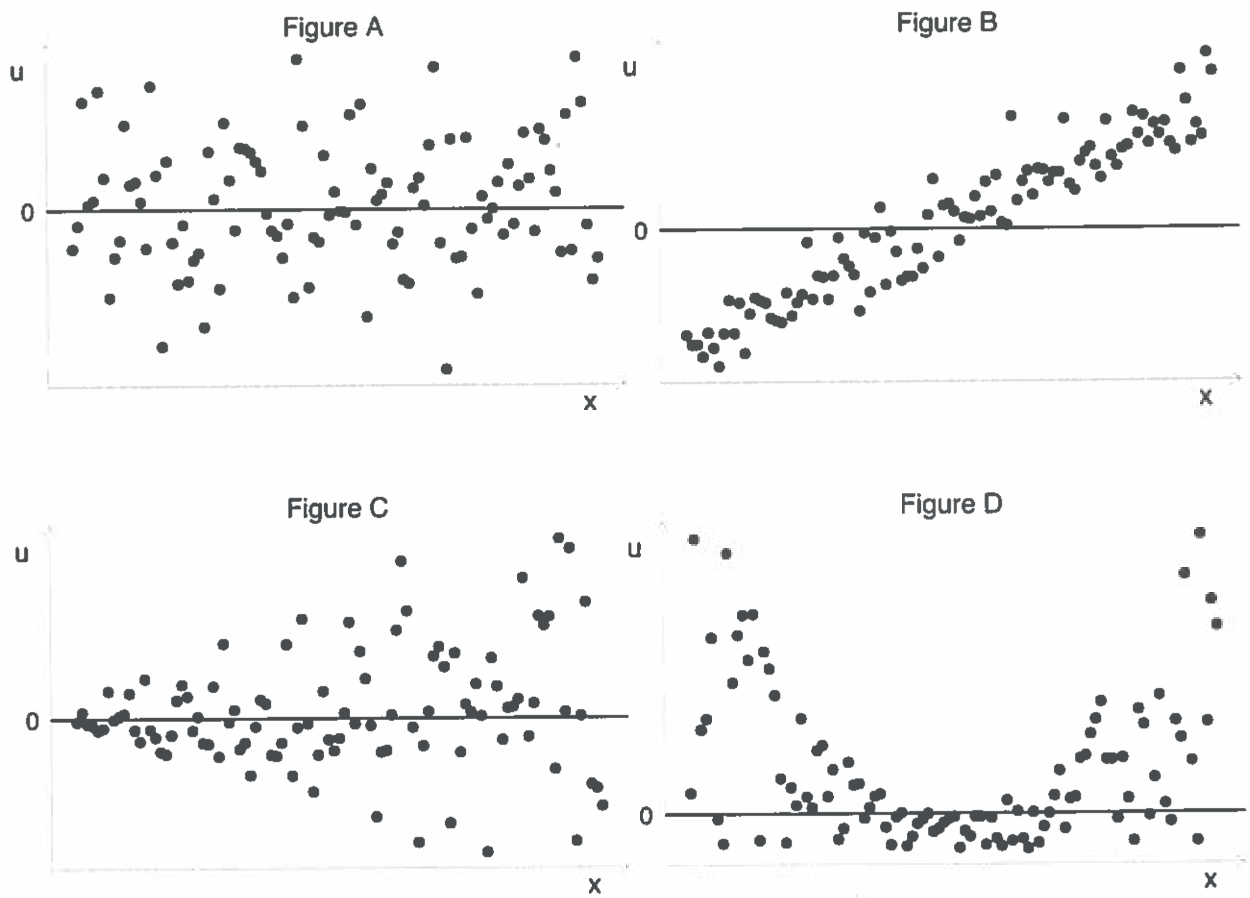


Figure A: (1pt) No violation: No trend in u vs x.

Figure B: (1pt) Violation: u increases w/. x.

Figure C: (1pt) No violation: E[u] does not change w/. x.

Figure D: (1pt) Violation: The mean of u changes w/. x.