

EC 421

**Midterm Review Questions**

07 February 2019

**Note:** These questions should help you review, along with the two problem sets. They are not necessarily comprehensive. I still suggest reviewing the problem sets and notes.

1. What is the difference between  $u_i$  and  $e_i$ ?
2. Why do we care about  $u_i^2$ ?
3. Explain each of our assumptions *in words*.
4. Which assumption does heteroskedasticity violate?
5. Which assumption does omitted-variable bias violate?
6. Load the `dplyr` package. You now have a dataset called `starwars`.
  - Regress the variable `mass` on the variable `height`. Conduct a *t* test and interpret the coefficient.
  - Regress the log of the variable `mass` on the variable `height`. Interpret the coefficient.
  - Regress the log of the variable `mass` on the log of the variable `height`. Interpret the coefficient.
  - For the linear-linear regression of `mass` on `height`, conduct a Breusch-Pagan test for heteroskedasticity.
  - For the linear-linear regression of `mass` on `height`, conduct a White test for heteroskedasticity.
  - Describe the steps you would need to run a Goldfeld-Quandt test for heteroskedasticity.
7. You are concerned about heteroskedasticity in a dataset. Following the Goldfeld-Quandt procedure, you calculate  $SSE_1=100$  and  $SSE_2=300$  (each group has 50 observations, and we have a simple linear regression model). Finish the Goldfeld-Quandt test for heteroskedasticity.
8. Is OLS biased or unbiased in the presence of heteroskedasticity? Is it still the 'best' linear unbiased estimator?
9. Draw two pictures of disturbances: (1) homoskedastic disturbances and (2) heteroskedastic disturbances. Be sure to label your axes.

10. You think the data underlying your econometric model may be heteroskedastic.
- What are your options?
  - What would you recommend to someone in this situation?
11. You have detected heteroskedasticity in your data/model.
- What are your options?
  - What happens if you don't do anything to deal with the heteroskedasticity?
12. How can misspecification lead to heteroskedasticity?
13. Weighted least squares (WLS) essentially divides observations by the standard deviation of their disturbance (*i.e.*, dividing by  $\sigma_i$ ). Explain the intuition for how this can increase efficiency.
14. If OLS is unbiased for our coefficients, why do we care about heteroskedasticity?
15. For the White, heteroskedasticity-robust standard error estimator, how do we estimate the coefficients?
16. What is the expected value of the estimator  $\mathbf{X}_1$ , *i.e.*, the value of the first observation? What is its variance?
17. What is required for an estimator to be consistent?
18. Can an estimator be unbiased and inconsistent? What about consistent and biased?
19. In the regression  $\mathbf{Income}_i = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{Education}_i + e_i$ , we omitted the variable **Ability**. Will our estimate  $\hat{\beta}_1$  (the effect of education on income) overestimate or underestimate the true value of  $\beta_1$ ? Explain.
20. Does omitted-variable bias cause OLS to be inconsistent?
21. How does a mis-measured explanatory variable affect OLS's estimates for the coefficients?
22. Does measurement error in the outcome variable matter? Explain.

31. What do we mean by *causality*?

32. Why is causality important? Are there instances where correlation is also important/interesting?

33. What do we mean by *prediction*? Does causality matter for prediction?

34. Write down the model that each of the lines of R code estimates. How would you interpret the coefficients in each model?

- `lm(y ~ x1 + x2)`
- `lm(y ~ x1 + x2 + x1:x2)`
- `lm(y ~ x1 + I(x1^2))`
- `lm(log(y) ~ x1 + x2)`
- `lm(log(y) ~ log(x1))`