EC 421

Midterm (5/6) Review Questions

Spring 2021

About these review problems:

- **We will not be providing answers to these questions.**
- These questions should help you review, along with the two problem sets. They are not 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 the assumptions required for OLS *in words*.

4. Which assumption does omitted-variable bias violate?

5. Load the `datasets` package. You now have a dataset called `attitude`. It measures a rating given by employees to a company. We'll use this

- Regress the variable `learning` (opportunities to learn) on the variable `rating` (rating given). Conduct a *t* test and interpret the coefficient.
- Regress the log of the variable `learning` on the variable `rating`. Interpret the coefficient.
- Regress the log of the variable `raises` (raises based on performance) on the log of the variable `rating`. Interpret the coefficient.
- For the linear-linear regression of `raises` on `rating`, conduct a Breusch-Pagan test for heteroskedasticity.
- For the linear-linear regression of `learning` on `rating`, conduct a White test for heteroskedasticity, describe the test.
- Describe the steps you would need to run a Goldfeld-Quandt test for heteroskedasticity.

6. You are concerned about heteroskedasticity in a dataset. Following the Goldfeld-Quandt procedure, you calculate $SSE_1=10$ and $SSE_2=15$ (each group has 14 observations, and we have a simple linear regerssion model). Finish the Goldfeld-Quandt test for heteroskedasticity.

7. Are OLS' standard errors biased or unbiased in the presence of heteroskedasticity? Is OLS still the 'best' linear unbiased estimator?

8. You think the data underlying your econometric model may be heteroskedastic.

- What are your options?
- What would you recommend to someone in this situation?

9. 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?

10. Explain why misspecification of your model can sometimes lead to heteroskedasticity?

11. Weighted least squares (WLS) essentially divides observations by the standard deviation of their disturbance (*i.e.*, dividing by σ_i). Explain the intuition for how this can increase efficiency.

12. fGLS which estimates the weights for a WLS regression is said to be 'biased, but consistent given the correct specification.' In your own words, explain what this means.

13. If OLS is unbiased for our coefficients, why do we care about heteroskedasticity?

14. For the White, heteroskedasticity-robust standard error estimator, how do we estimate the coefficients?

15. What is the expected value of the estimator $\dfrac{\sum_i^N X_i}{\frac{N}{2}}$? What is its variance?

16. What is required for an estimator to be consistent?

17. Can an estimator by unbiased and inconsistent? What about consistent and biased? In your own words, explain what is the best case scenario for an estimator?

18. In the regression $\mathbf{Income}_i = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{Education}_i + e_i$, we omitted the variable $\mathbf{Ability}$. Will our estimate $\hat{\beta}_1$ (the effect of education on income) overestimate or underestimate the true value of β_1? Explain.

19. Does omitted-variable bias cause OLS to be inconsistent?

20. How does a mis-measured explanatory variable affect OLS's estimates for the coefficients?

21. Does measurement error in the outcome variable matter? Explain.

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

23. Is the OLS regression estimating the coefficient β_1 in the model
$$\text{Cig}_t = \beta_0 + \beta_1 \text{Cig}_{t-1} + \beta_2 \text{income\_t} + u_i$$ biased? Explain why.

24. In the standard regression model of average income on log housing prices, write down the ADL(2,1) specification

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

26. How do time-series data differ from cross-sectional data?

27. What is the difference between a **static** time-series model and a **dynamic** time-series model?

28. 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))`