EC 421
Midterm

12 February 2019

Full Name ←

UO ID ←

No phones, calculators, or outside materials.

A. True/False and Multiple Choice

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 the coefficient β_1 as the expected percentage change in y_i due to a 1-percent increase in x_i.

02. **[T/F]** (**2pts**) The difference between the White test for heteroskedasticity and the Breusch-Pagan test for heteroskedasticity is that the Bruesch-Pagan test uses interactions and squared terms.

03. **[T/F]** (**2pts**) If the *p*-value corresponding to our estimate of β_1 is 0.08, then we reject the null hypothesis at the 5-percent level.

04. **[T/F]** (**2pts**) Heteroskedasticity occurs when $E[u_i|x_i] \neq 0$.

05. **[T/F]** (**2pts**) Omitted-variable bias results in OLS estimates that are biased toward zero.

06. **[T/F]** (**2pts**) If we have an omitted variable, we can use weighted least squares (WLS) to avoid bias.

07. **[T/F]** (**2pts**) Exogeneity requires that the mean of the disturbances ($E[u_i]$) is uncorrelated with all explanatory variables (x_i).

08. **[T/F]** (**2pts**) If you omit a variable from your regression, then you will have omitted-variable bias.

09. **[T/F]** (**2pts**) OLS's consistency is also affected by omitted-variable "bias".

10. [T/F] (**2pts**) In the presence of heteroskedasticity, OLS estimates for the coefficients are biased.

11. [T/F] (**2pts**) In the presence of heteroskedasticity, OLS estimates for the standard errors are biased.

12. [T/F] (**2pts**) If an estimator is biased, then it is also inconsistent.

13. [T/F] (**2pts**) Weighted least squares (WLS) gives more weight to observations with low-variance disturbances and less weight to observations with high-variance disturbances.

14. Consider the model $\text{Employed}_i = \beta_0 + \beta_1 \text{School}_i + \beta_2 \text{Female}_i + u_i$, where Employed_i is a binary indicator for whether individual i is employed, School_i gives the number of years of schooling for individual i, and Female_i is a binary indicator for whether i is female.

a. [T/F] (**2pts**) The coefficient β_1 gives the expected increase in the probability of employment (in *percentage points*) for a one-year increase in schooling (holding everything else constant).

b. [T/F] (**2pts**) This model allows the effects of schooling to vary by gender.

c. [T/F] (**2pts**) If race is correlated with education and does not affect employment status, then OLS estimates of β_1 will be biased.

15. Multiple choice (**4pts**) Choose **all** correct answers.

If an estimator $\hat{\theta}$ is unbiased for θ, then

A. $\hat{\theta} = \theta$ **B.** $E\left[\hat{\theta}\right] = \theta$ **C.** $\text{plim}\left(\hat{\theta}\right) = \theta$ **D.** $E\left[\hat{\theta}\right] - \theta = 0$

16. Multiple choice (**4pts**) Choose **all** correct answers.

Which of the following are part of our assumptions?

A. $E[u_i|x_i] = 0$ **B.** $\text{Var}(x) = 0$ **C.** $E[u_i] = 0$ **D.** $\text{Var}(u_i) = 0$

B. Short Answer

60 points

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

17. (**3pts**) Define what we mean by "the standard error of an estimator".

18. (**3pts**) What does it mean if the estimator $\hat{\beta}$ is consistent for β?

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

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{a}$$
$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \tag{b}$$

20. (**2pts**) What is measurement error?

21. (**2pts**) How does measurement error in an explanatory variable affect OLS estimates?

22. For the model $\text{Health}_i = \beta_0 + \beta_1 \text{Income}_i + u_i$

a. (**4pts**) Which **two** conditions must be true for an omitted variable to bias our estimates of β_1?

b. (**2pts**) Provide an example of an omitted variable that could cause bias in this scenario. Explain your reasoning.

23. For the model $\textbf{Income}_i = \beta_0 + \beta_1 \textbf{Female}_i + u_i$, where $\textbf{Female}_i$ is a binary indicator for whether individual i is female,

a. (**3pts**) The term __________ gives the average income for men.

b. (**3pts**) The term __________ gives the difference between the average income for women and the average income for men.

c. (**3pts**) The sum ____________________ gives the average income for women.

24. In our proof of the consistency of the OLS estimator for β_1 (for simple linear regression), we showed

$$\text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x,\,u)}{\text{Var}(x)}$$

a. (**2pts**) If the OLS estimator $\hat{\beta}_1$ is consistent for β_1, then what does the right-hand side of this equation equal?

b. (**3pts**) What is the "next step" (and last step) in this derivation? How/why do we get the result that OLS is indeed consistent for β_1?

25. (**4pts**) Compare **and** contrast the concepts of consistency and unbiasedness.
Hint: "Compare" means how they are similar; "contrast" means how they differ.

26. Time for some drawing.

a. (**3pts**) For $y_i = \beta_0 + \beta_1 x_i + u_i$, draw a picture/plot that depicts **homoskedastic** disturbances. Make sure you label both axes. Add an explanation if you'd like.

b. (**3pts**) For $y_i = \beta_0 + \beta_1 x_i + u_i$, draw a picture/plot that depicts **heteroskedastic** disturbances. Make sure you label both axes. Add an explanation if you'd like.

27. Suppose we model the relationship between crime and policing at the city level using

$$\text{Crime}_i = \beta_0 + \beta_1 \text{Police} + u_i \tag{2}$$

where i indexes a city, Crime_i gives the number of crimes committed in city i, and Police_i gives the number of police officers working in city i.

a. (**2pts**) We estimate $\hat{\beta}_1 = -3.1$. How do we interpret this coefficient?

b. (**2pts**) We estimate $\hat{\beta}_0 = 537.3$. How do we interpret this coefficient? Explain why this interpretation doesn't really make sense.

c. (**5pts**) We're concerned about heteroskedasticity and decide to run a White test. Write out the steps we need to carry out to conduct a White test, describing each step (including any hypotheses, regression equations, *etc.*).

d. (**4pts**) Suppose we ran a White test and calculated an LM test statistics of 7.3, which implies a *p*-value of 0.026 (using a χ^2 with 2 degrees of freedom). What do we conclude from this test statistic and *p*-value? Include an interpretation.

e. (**2pts**) Which part of the White-test procedure that you outlined in part (**c.**) changes if we opt for a Breusch-Pagan test (as opposed to a White test)? What is the change?

f. (**2pts**) We are also concerned about omitted-variable bias—specifically with respect to a city's average income (Income_i). In class we showed that the probability limit of the OLS estimator for β_1 is

$$\text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(\text{Police, Income})}{\text{Var}(\text{Police})} \tag{2}$$

If (*i*) cities with higher incomes have more police officers and (*ii*) higher incomes lead to less crime, then *how* (which direction) will OLS be biased for β_1 in equation (2)? Explain your answer.

C. Extra Credit

8 points

EC_1 **[T/F]** (**2pts**) Omitted-variable bias has nothing to do with whether we interpret regression estimates as causal.

EC_2 (**2pts**) Write down the regression equation that we would estimate in the following line of `R` code (*i.e.*, the equation with βs).

```
lm(crime ~ police + income + police:income, data = city_df)
```

EC_3 (**2pts**) Draw a plot of disturbances that depicts a **violation of exogeneity**.

EC_4 (**2pts**) Draw a plot of **heteroskedastic disturbances** for which the Goldfeld-Quandt test would fail to find significant evidence of heteroskedasticity.