EC 421
Midterm

12 February 2019

Full Name ← KEY

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.

F

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.

F

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.

F

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

F

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

F

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

F

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

T

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

F

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

T

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

F

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

T

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

F

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.

T

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).

T

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

F

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.

F

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.** $E\left[\hat{\theta}\right] - \theta = 0$

B, D

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.** $\text{Var}(u_i) = 0$

A, C

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".

Std. dev. of estimator's distribution.

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

Options:

A. $\text{plim}\ \hat{\beta} = \beta$ (the prob. limit of $\hat{\beta}$ is β)

B. $\hat{\beta}$ converges (in probability) to β

C. As $n \to \infty$ (sample size goes to infinity), the estimator's distribution collapses to a point at β

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

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad \text{(a)}$$
$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \text{(b)}$$

u_i is an unknown population term that describes how far individual i is from the population trend line. (u_i is is called the disturbance.)

e_i is the sample-based error/residual that tells us how far individual i sits from our regression line.

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

Measurement error is a situation in which one of our variables is inaccurately measured (mismeasured).

Also: Measurement error is when our variable is measured with noise.

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

Measurement error (in the expl. variable) attenuates OLS estimates, i.e., it biases our estimates toward zero.

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?

1. The omitted variable must affect Health.
2. The omitted variable must be correlated with Income.

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

Type of diet, exercise, age, air quality, ...

For age: Age likely affects Health and correlates with income.

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

a. (**3pts**) The term β_0 gives the average income for men.

b. (**3pts**) The term β_1 gives the difference between the average income for women and the average income for men.

c. (**3pts**) The sum $\beta_0 + \beta_1$ 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?

β_1

Accept: $\frac{Cov(x,u)}{Var(x)} = 0$

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?

$Cov(x,u) = 0$ by assumption (exogeneity)

Not needed: [Also: $Var(x) \neq 0$]

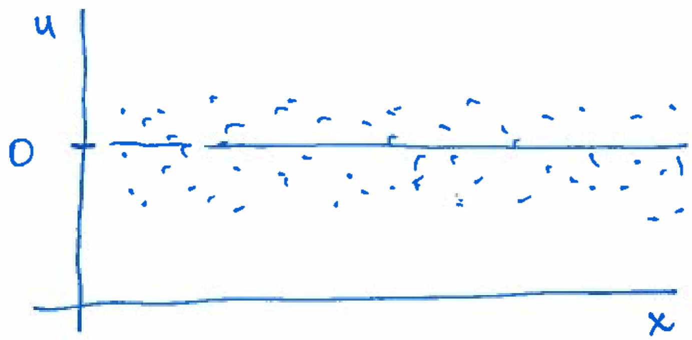
25. (**4pts**) Compare **and** contrast the concepts of consistency and unbiasedness.
Hint: "Compare" means how they are similar; "contrast" means how they differ.

Both terms describe some type of behavior for an estimator and its distribution.

The difference: consistency describes how the estimator behaves as the sample size (n) approaches ∞. Unbiasedness only asks about the expected value (mean) of an estimator's distribution for a fixed sample size.

26. Time for some drawing.

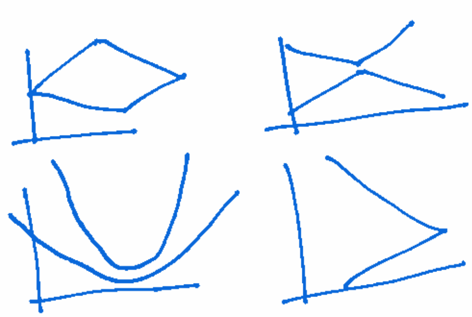
a. (**3pts**) For $y_i = \beta_0 + \beta_1 x_i + u_i$, draw a picture/plot that depicts **homoskedastic** disturbances. Make sure you label both axes. Add an explanation if you'd like.



Not necessary to center on zero

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.

other shapes :



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

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

where i indexes a city, $\textbf{Crime}_i$ gives the number of crimes committed in city i, and $\textbf{Police}_i$ 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?

For each additional police officer, we expect the number of crimes to decrease by 3.1.

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

If a city has zero police officers, we expect there to be 537.3 crimes.

This interpretation is a bit strange because we don't expect there to be cities without police.

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.*).

1pt (1.) Estimate via OLS: $\text{Crime}_i = \beta_0 + \beta_1 \text{Police}_i + u_i$

2pt (2.) Grab residuals from step (1) and run regression:

$$e_i^2 = \alpha_0 + \alpha_1 \text{Police}_i + \alpha_2 \text{Police}_i^2 + v_i$$

2pt (3.) Hypothesis test of

H_0: $\alpha_1 = \alpha_2 = 0$
or ($\alpha_1 = 0$ and $\alpha_2 = 0$)
or (Homoskedasticity)
or ($\sigma_i^2 = \sigma_j^2$ for all i, j)

vs

H_A: $\alpha_1 \neq 0$ or $\alpha_2 \neq$
or H_A: heteroskedasticity
or H_A: $\sigma_i^2 \neq \sigma_j^2$ for some i, j

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.

2pt We reject the null hypothesis (H_0).

2pt We conclude there is statistically significant evidence of heteroskedasticity at the 5% level.

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?

We would drop $\alpha_2 \text{Police}_i^2$ from our regression with the squared residuals.

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

$\beta_2 \leftarrow$ effect of income on crime

Typo

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

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.

* Due to typo, everyone gets the 2pts.

If they get the answer correct +2pts (extra credit).

In this scenario, $\text{plim}\,\hat{\beta}_1 = \beta_1 + (-)\dfrac{(+)}{(+)}$

$(+) \leftarrow$ policing increases w/. income

$(+) \leftarrow$ Var > 0

$(-) \leftarrow$ effect of income on crime.

$= \beta_1 + (-)$

which means we will be biased downward (under-estimating the effect of policing on crime).

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.

F

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)
```

$$\text{Crime}_i = \beta_0 + \beta_1 \text{police}_i + \beta_2 \text{income}_i + \beta_3 \text{income}_i \times \text{police}_i + u_i$$

EC$_3$ (2pts) Draw a plot of disturbances that depicts a **violation of exogeneity**.

Two examples...

or

EC$_4$ (2pts) Draw a plot of **heteroskedastic disturbances** for which the Goldfeld-Quandt test would fail to find significant evidence of heteroskedasticity.