Name: _____

# In-class final, EC421

140 points possible

## 1 True or false (75 points; 30 questions)

**Note** In this section, select the correct answer (true or false). You do not need to explain your answer.

1. (2.5 points) **[T/F]** If you omit a variable that (1) affects your outcome and (2) correlates with an included variable, least-squares regression will be biased for estimates of the coefficients.
2. (2.5 points) **[T/F]** If OLS is biased by an omitted variable, it can still be consistent.
3. (2.5 points) **[T/F]** In the model $\text{Health}_t = \beta_0 + \beta_1 \text{Pollution}_t + \beta_2 \text{Pollution}_{t-1} + u_t$, health only depends upon contemporaneous pollution.
4. (2.5 points) **[T/F]** When disturbances' means differ across groups, we have heteroskedasticity.
5. (2.5 points) **[T/F]** Adding additional variables mechanically increases $R^2$.
6. (2.5 points) **[T/F]** If an estimator is biased, it will also be inconsistent.
7. (2.5 points) **[T/F]** Heteroskedastic disturbances make OLS biased for estimating coefficients.
8. (2.5 points) **[T/F]** If we estimate the econometric model below via regression, $\widehat{\beta}_1$ will equal average(education for females) minus average(education for non-females).

$$\text{Education}_i = \beta_0 + \beta_1 \text{Female}_i + u_i$$

9. (2.5 points) **[T/F]** Exogeneity essentially says that the disturbance must be independent of your explanatory variables.
10. (2.5 points) **[T/F]** If OLS is biased for estimating $\beta_1$ below, then it is also inconsistent for estimating $\beta_1$.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + u_t$$

11. (2.5 points) **[T/F]** A $p$-value of 0.97 suggests the data do not support the null hypothesis.

12. (2.5 points) **[T/F]** Random walks are stationary.

13. (2.5 points) **[T/F]** In the Rubin causal model, $y_{1i}$ refers to the outcome for individual $i$ when it does not receive treatment.

14. (2.5 points) **[T/F]** If $u_i$ correlates with $u_j$, then exogeneity is violated.

15. (2.5 points) **[T/F]** In the presence of heteroskedasticity, WLS can be more efficient than OLS.

16. (2.5 points) **[T/F]** If a Goldfeld-Quandt test finds that $\text{SSE}_1$ equals $\text{SSE}_2$, then it will conclude that there is statistically significant evidence of heteroskedasticity.

17. (2.5 points) **[T/F]** Measurement error in an explanatory variable tends to cause OLS to underestimate the true effect of that variable on the outcome.

18. (2.5 points) **[T/F]** Dynamic models with lagged outcome variables always violate contemporaneous exogeneity.

19. (2.5 points) **[T/F]** The econometric model below allows the effect of gender on education to depend upon the individual's age.

$$\text{Education}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i \times \text{Age}_i + u_i$$

20. (2.5 points) **[T/F]** Correlated disturbances make OLS biased when estimating standard errors.

21. (2.5 points) **[T/F]** Randomizing the explanatory variable helps avoid selection bias.

22. (2.5 points) **[T/F]** Weighted least squares (WLS) upweights individuals with high-variance disturbances and downweights individuals with low-variance disturbances.

23. (2.5 points) **[T/F]** A variable will cause omitted-variable bias in OLS estimates for coefficients when the following things are true:

    1. A variable is omitted from the regression.

    2. The omitted variable correlates with one of the included regressors.

24. (2.5 points) **[T/F]** The heteroskedasticity-robust standard error estimator is unbiased when the disturbance is homoskedastic.

25. (2.5 points) **[T/F]** OLS is biased when estimating the model below.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 y_{t-1} + u$$

26. (2.5 points) **[T/F]** In the model

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

If $\widehat{\beta}_1 = 0.57$, then, on average, a one-percent increase in education generates a 57-percent increase in income (all else equal).

27. (2.5 points) **[T/F]** If $\text{Var}(u_t) = 4$ for $t$ in $\{1, \dots, 10\}$ and $\text{Var}(u_t) = 5$ for $t$ in $\{11, \dots, 20\}$, then $u_t$ is variance stationary.

28. (2.5 points) **[T/F]** If the variable $x_t$ is a random walk, then $x_t - x_{t-1}$ is nonstationary.

29. (2.5 points) **[T/F]** An autocorrelated disturbance biases OLS when estimated the coefficients.

30. (2.5 points) **[T/F]** The fundamental problem of causal inference is that a comparison of the treatment group to the control group tends to be biased by selection bias.

## 2 Short answer (65 points; 10 questions)

**Note** In this section, briefly answer the questions/prompts in 1–3 short (and complete) sentences. We will deduct points for excessively long answers.

31. (5 points) Define the concept of a *standard error*.

32. (12 points) For each of the six lettered entries in the table below (a–f), answer whether OLS is (1) unbiased or (2) consistent when estimating the coefficients in a linear regression model. For example, your answer *for each letter* should look something like "biased and consistent" or "unbiased and inconsistent."

You do not need to explain your answers.

| | | *Dynamic models* | |
|---|---|---|---|
| | Static model | Lagged expl. vars. | Lagged outcome var. |
| $u_t$ is **not** autocorrelated | (a) | (b) | (c) |
| $u_t$ is autocorrelated | (d) | (e) | (f) |

(a) Unbiased or consistent?

(b) Unbiased or consistent?

(c) Unbiased or consistent?

(d) Unbiased or consistent?

(e) Unbiased or consistent?

(f) Unbiased or consistent?

33. (5 points) Explain why we might expect the following model's disturbance to be autocorrelated. Include at least one specific example in your answer.

$$\text{Health}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$$

34. (5 points) Explain how randomized experiments avoid omitted-variable bias.

35. (5 points) What issues can *non-stationary* data cause? Explain your answer.

36. (5 points) Should we be concerned about autocorrelation in a cross-sectional dataset? Explain.

37. (5 points) Compare and contrast *probability limits* and *expected values*.

38. (5 points) For the model below, suppose our estimates of the parameters are $\hat{\beta}_0 = 7.4$, $\hat{\beta}_1 = -0.02$, and $\hat{\beta}_2 = 0.01$. Interpret each of the coefficient estimates (you can ignore the intercept).

$$\log \text{Crime}_t = \beta_0 + \beta_1 \text{Police}_t + \beta_2 \text{Population}_t + u_t$$

39. For the model below, assume the variable $\text{Youth}_i$ is an indicator that equals one when the individual is less than 16 years old. Health and pollution are continuous variables.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Youth}_i + \beta_2 \text{Pollution}_i + \beta_3 \text{Youth}_i \times \text{Pollution}_i + u_i$$

(a) (3 points) What is the expected health for a youth exposed to "50 units" of pollution?

(b) (3 points) What is the effect of an additional unit of pollution for youth?

(c) (3 points) Explain how the regression above estimates two lines—and provide the intercepts and slopes of the two lines.

40. Suppose we are analyzing the following model of Oregon's (annual) GDP and state income tax

$$\text{GDP}_t = \beta_0 + \beta_1(\text{Income Tax})_t + \beta_2(\text{Income Tax})_{t-1} + \beta_3(\text{Income Tax})_{t-2} + u_t$$

(a) (3 points) What is the effect of this year's income tax on next year's GDP?

(b) (3 points) What is the total effect of income tax on GDP?

(c) (3 points) Should we include a lagged outcome variable? Explain your answer.