Name: _____

In-class final, EC421

Part 1: True or false (60 points)

Note: In this section, select the correct answer. You do not need to explain your answer.

1. **[T/F]** (2pts) For the model below: The effect of experience depends on the level of education.

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Experience}_i + \beta_2 \text{Education}_i + \beta_3 \text{Experience}_i \times \text{Education}_i + u_i$$

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

3. **[T/F]** (2pts) When the disturbance is *not* autocorrelated, OLS will be biased for estimating the coefficients in the following model:

$$\text{Crime}_t = \beta_0 + \beta_1 \text{Inflation}_t + \beta_2 \text{Inflation}_{t-1} + \beta_3 \text{Inflation}_{t-2} + u_t$$

4. **[T/F]** (2pts) In the two models below: β_1 has the same interpretation.

$$y_t = \beta_0 + \beta_1 x_t + u_t$$
$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \Delta u_t$$

5. **[T/F]** (2pts) The variable x_t—as defined below—is nonstationary.

$$x_t = x_{t-1} + t + \varepsilon_t$$

   where ε_t is 'well behaved' (stationary).

6. **[T/F]** (2pts) Measurement error in Health_i will bias estimates of β_1 toward zero (see below).

$$\text{Health}_i = \beta_0 + \beta_1 \text{Pollution}_i + u_i$$

7. **[T/F]** (2pts) To consistently estimate the effect of x on y, instrumental variables requires that x is exogenous.

8. **[T/F]** (2pts) The model below implicitly includes many lags of inflation.

$$\text{Crime}_t = \beta_0 + \beta_1 \text{Inflation}_t + \beta_2 \text{Crime}_{t-1} + u_t$$

9. **[T/F]** (2pts) For an instrument to be exogenous, it can only affect the outcome through the endogenous variable.

10. **[T/F]** (2pts) If the disturbances of individual i and individual j are correlated, then OLS is biased for the coefficients in

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

11. **[T/F]** (2pts) If the disturbances of individual i and individual j are correlated, then OLS is biased for the standard errors of the coefficients in

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

12. **[T/F]** (2pts) Exogeneity requires that the regressors are entirely unrelated to the disturbance.

13. **[T/F]** (2pts) If our explanatory variable is exogenous, then we can interpret its effect on the outcome as causal.

14. **[T/F]** (2pts) In the Neyman-Rubin framework, $y_{1,i}$ describes individual i's outcome when she receives treatment.

15. **[T/F]** (2pts) If the estimator $\hat{\alpha}$ is consistent for the parameter α, then $E[\hat{\alpha}] = \alpha$.

16. **[T/F]** (2pts) OLS is always biased for dynamic models with lagged outcome variables.

17. **[T/F]** (2pts) For the model below, if $\hat{\beta}_1 = 0.3$, then we can interpret the estimated coefficient as a one-unit increase in x tends to increase y by 30% (all else equal).

$$\log(y_t) = \beta_0 + \beta_1 \log(x_t) + u_t$$

18. **[T/F]** (2pts) In practice (in the "real world"), you should generally avoid standard errors that assume homoskedasticty.

19. **[T/F]** (2pts) The "LS" in OLS means that the estimates for the βs minimize SSE.

20. **[T/F]** (2pts) OLS's linearity requirement prohibits nonlinear transformations like x^2.

21. **[T/F]** (2pts) The following disturbance violates mean stationarity

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where $u_0 = 0$ and $E[\varepsilon_t] = 0$.

22. **[T/F]** (2pts) If time-series models satisfy contemporaneous exogeneity, then they are unbiased.

23. **[T/F]** (2pts) Adding additional variables does not always increase R^2.

24. **[T/F]** (2pts) The relevance requirement of instrumental variables is untestable.

25. **[T/F]** (2pts) If we omit a variable that is positively correlated with a regressor and has a negative effect on our outcome, then OLS will be biased downward.

26. **[T/F]** (2pts) The first stage in two-stage least squares (2SLS) and instrumental variables (IV) is when we regress the outcome variable on the instrumental variable.

27. **[T/F]** (2pts) 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.

28. **[T/F]** (2pts) Dynamic models violate contemporaneous exogeneity.

29. **[T/F]** (2pts) In the following model, β_1 gives the average difference between college graduates and non-college-graduates.

$$\text{Wage}_i = \beta_0 + \beta_1(\text{College Graduate})_i + u_i$$

30. **[T/F]** (2pts) Homoskedasticity requires $\text{Var}(u_i) = \sigma^2$ for all i.

Part 2: Short answer (50 points)

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. (5pts) How is the standard error useful in learning about the uncertainty in an estimator? (Do not just give the definition of the standard error.)

32. (5pts) Compare and contrast consistency and unbiasedness

33. (5pts) Use math to show why random walks are considered 'difference stationary'.

34. (5pts) Define the two main requirements for a valid instrument—explaining each requirement.

35. (5pts) Suppose I want to estimate the effect of college graduation on later-life income. Would *parents' college graduation status* be a valid instrument? Explain your answer.

36. (5pts) Explain why we might expect the following model's disturbance to be autocorrelated. Give specific examples.

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

37. (5pts) Explain how randomized experiments avoid selection bias.

For the next three problems, use the model below.

$$\text{Wage}_i = \beta_0 + \beta_1(\text{College Graduate})_i + \beta_2\text{Female}_i + \beta_3(\text{College Graduate})_i \times \text{Female}_i + u_i$$

38. (5pts) Which combination of parameters gives the expected wage of a female college graduate?

39. (5pts) Which combination of parameters gives the expected wage of non-female non-college-graduates?

40. (5pts) Which combination of the parameters gives the expected wage difference between non-female non-college graduates and female non-college graduates?

Part 3: Long-ish answer (10 points)

Note: Answer the prompt with 1–3 paragraphs.

36. (10 pts) Suppose you work in analytics for a company that sells video games. Your boss gives you monthly data from the company's last 5 years. The data include (a) the month and year, (b) the number of games sold, and (c) the amount of money the company spent on advertising.

    Your boss wants to know how successful advertising has been in increasing sales.

    Write down the 'best' model you could estimate with these data and then describe the assumptions/requirements for the model to be unbiased and/or consistent. What would you say/do?