Name: _____

In-class final, EC421

180 points possible

1 True or false (105 points; 35 questions)

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

1. (3 points) **[T/F]** Even if an estimator is biased, it can still be consistent.
2. (3 points) **[T/F]** The variable $x_t = 0.3t + \varepsilon_t$ is stationary (assume ε_t is mean zero).
3. (3 points) **[T/F]** Exogeneity requires that the regressors are entirely unrelated to the disturbance.
4. (3 points) **[T/F]** Measurement error in an explanatory variable tends to bias OLS's coefficient estimates toward zero.
5. (3 points) **[T/F]** The instrumental variable (IV) estimator requires that the regressor of interest is exogenous.
6. (3 points) **[T/F]** When disturbances' means differ across groups, we have heteroskedasticity.
7. (3 points) **[T/F]** In the Neyman-Rubin framework, Avg$(y_1|D_i = 0)$ represents the average untreated outcome for the treated group.
8. (3 points) **[T/F]** Adding additional variables does not always increase R^2.
9. (3 points) **[T/F]** For the model

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

   income's total effect of income on health is $\beta_1 + \beta_2 + \beta_3$.
10. (3 points) **[T/F]** Heteroskedastic disturbances make OLS biased for estimating coefficients.
11. (3 points) **[T/F]** The presence of omitted-variable bias implies that OLS is also inconsistent.

12. (3 points) **[T/F]** By including a lagged dependent variable, we implicitly include lags of explanatory variables.

13. (3 points) **[T/F]** If the disturbance is heteroskedastic, the usual t and F tests are not valid.

14. (3 points) **[T/F]** If we estimate the econometric model below via regression, $\hat{\beta}_1$ will equal Avg(Health | College Graduate) − Avg(Health | Non-Graduate).

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

15. (3 points) **[T/F]** Exogeneity essentially says that the disturbance must be independent of your explanatory variables.

16. (3 points) **[T/F]** OLS is biased when estimating coefficients in models with lagged outcome variables.

17. (3 points) **[T/F]** The first stage of two-stage least squares (2SLS) involves regressing the outcome variable on the instrumental variable.

18. (3 points) **[T/F]** If OLS is biased for estimating β_1 below, then it is also inconsistent for estimating β_1.

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

19. (3 points) **[T/F]** A p-value of 0.999 suggests the data do not support the null hypothesis.

20. (3 points) **[T/F]** Random walks are non-stationary.

21. (3 points) **[T/F]** In the Rubin causal model, y_{1i} refers to the outcome for individual i when it does not receive treatment.

22. (3 points) **[T/F]** If u_i correlates with u_j, then exogeneity is violated.

23. (3 points) **[T/F]** If a Goldfeld-Quandt test finds that SSE_1 equals SSE_2, then it will conclude that there is statistically significant evidence of heteroskedasticity.

24. (3 points) **[T/F]** In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$, the parameter β_1 gives the effect of income in the previous time period on births in the current period.

25. (3 points) **[T/F]** Dynamic models with lagged outcome variables always violate contemporaneous exogeneity.

26. (3 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.

27. (3 points) **[T/F]** The econometric model below allows the effect of gender on education to depend upon the individual's age.

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

28. (3 points) **[T/F]** Correlated disturbances make OLS biased when estimating standard errors.

29. (3 points) **[T/F]** Randomizing the explanatory variable helps avoid selection bias.

30. (3 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.

31. (3 points) **[T/F]** The outcome y_1 is unobserved in the control group.

32. (3 points) **[T/F]** In the model below, if the disturbance is *not* autocorrelated, then OLS consistently estimates the coefficients.

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

33. (3 points) **[T/F]** In the model

$$\log \text{Quantity}_t = \beta_0 + \beta_1 \log \text{Price}_t + u_t$$

    if $\hat{\beta}_1 = -0.9$, then, on average, a one-percent increase in price generates a 0.9-percent reduction in quantity (all else equal).

34. (3 points) **[T/F]** If $\text{Var}(u_t) = 13 + 4t$, then u_t is variance stationary.

35. (3 points) **[T/F]** If the variable x_t is a random walk, then $x_t - x_{t-1}$ is nonstationary.

2 Short answer (75 points; 16 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.

36. Suppose you estimate the following model with OLS

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

and find parameter estimates $\hat{\beta}_0 = 30$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -5$, and $\hat{\beta}_3 = 0.5$.

(a) (5 points) Interpret $\hat{\beta}_3$ in the context of the model.

An additional year of education increases income 0.5 "units" more for females than for non-females.

(b) (5 points) What is the expected income for a non-female individual with 12 years of education?

$30 + 1 \times 12 + \cancel{0} + \cancel{0.5 \times 12 \times 0} = 30 + 12 = 42$

(c) (5 points) Explain how the regression above estimates two lines—and provide the slope of each line.

We have diff. intercepts and slopes for females and non-females.

The slopes:

- Non-females: β_1 (1)
- Females: $\beta_1 + \beta_3$ ($1 + 0.5 = 1.5$)

(d) (5 points) Explain how you could use this model to test for gender-based discrimination.

β_2 and β_3 are both helpful here.

↳ β_2 tells us how incomes of females and non-females differ at low levels of education. If $\beta_2 \neq 0$, then there is a gender diff. @ low lvls. of education. $\beta_2 < 0$ → females are "underpaid."

↳ $\beta_3 \neq 0$ tells us that females have diff. returns to educ. $\beta_3 < 0$ → less income per unit of educ. for females.

37. People commonly debate the impact of the minimum wage on unemployment. Suppose you have data on unemployment and minimum wage for 100 cities in the United States.

(a) (3 points) Write down the econometric (regression) model you would use to estimate the effect of minimum wage on unemployment.

$$\text{Unemployment}_i = \beta_0 + \beta_1 \text{Min.Wage}_i + u_i$$

(b) (2 points) Which parameter in your model gives the effect of minimum wage on unemployment?

β_1

(c) (5 points) Suppose you estimate your model (in (a) above) with OLS. Will these OLS estimates provide causal evidence about the effect of minimum wage on unemployment? Explain your answer.

Only if min. wage changes are exogenous, ie, if they occur independent of u_i (other determinants of unemployment).

(d) (5 points) Could instrumental variables help here? Explain.

Yes! If we have valid IV, we can recover the causal effect even if min. wage is not exogenous.

38. (5 points) Define the concept of a *standard error*.

SE: The standard deviation of an estimator's distribution.

39. (5 points) Using mathematical notation, explain the concept of *exogeneity*. Then explain why exogeneity is important for OLS.

$$\mathbb{E}[u_i \mid X] = 0$$

Importance: We want to attribute changes in X to changes in y, so we need u (other factors that affect y) to stay fixed (on avg).

40. (5 points) Suppose you are analyzing time-series data for your job. Why should you be concerned about non-stationarity?

Non-stationary data can lead to spurious results.

41. (5 points) The probability limit of the instrumental variables (IV) estimator (with instrument z, endogenous regressor x, and disturbance u) is

$$\text{plim}\ \hat{\beta}_{\text{IV}} = \beta + \frac{\text{Cov}(z,u)}{\text{Cov}(z,x)}$$

Use the equation above to explain how a valid instrument makes the IV estimator consistent.

We have 2 requirements:

1. Relevance: $\text{Cov}(x,z) \neq 0$
2. Exogeneity: $\text{Cov}(z,u) = 0$

$$\text{plim}\ \hat{\beta}_{IV} = \beta + \frac{0}{C} = \beta$$

42. (5 points) Explain how correlated disturbances can affect OLS.

They can bias OLS's SE estimates (and thus t-tests, Conf. intervals, etc.).

43. (5 points) Explain how randomized experiments avoid omitted-variable bias.

Randomized experiments generate exogenous variation, on avg. trt. and control are comparable (no selection bias).

44. (5 points) Suppose you want to know how a recent advertising campaign affected sales. You have monthly data on sales (in dollars) and advertising spending (in dollars) for several years. Write down the econometric model you would use to estimate the effect of advertising on sales.

$$Sales_t = \beta_0 + \beta_1 Advertising_t + \beta_2 Advertising_{t-1} + u_t$$

(could be static or ADL too)

45. (5 points) What are the key assumptions required to interpret your coefficient estimates from the model in the previous question as causal? Are these assumptions plausible in this context? Explain.

Accept stationarity too.

Causality requires exogeneity. If ad. campaign started "randomly", then we might be fine. If we started the campaign started in response to low sales, then we def. do not have exog.