Name: _____

In-class final, EC421

120 points possible

1 True or false (50 points)

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

1. (2.5 points) **[T/F]** In the presence of omitted-variable bias, ordinary least squares (OLS) regression is still consistent for estimating the βs.
2. (2.5 points) **[T/F]** When disturbances' means differ across groups, we have heteroskedasticity.
3. (2.5 points) **[T/F]** Adding additional variables mechanically increases R^2.
4. (2.5 points) **[T/F]** If an estimator is biased, it is also inconsistent.
5. (2.5 points) **[T/F]** Static time-series models include lags for the explanatory variables but not the dependent variable.
6. (2.5 points) **[T/F]** Random walks are non-stationary because their means change with time.
7. (2.5 points) **[T/F]** Randomizing a variable typically ensures exogeneity.
8. (2.5 points) **[T/F]** Correlated disturbances make OLS biased when estimating standard errors.
9. (2.5 points) **[T/F]** If we estimate the econometric model below via regression, $\hat{\beta}_1$ will equal Average(Income for females) $-$ Average(Income for non-females).

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

10. (2.5 points) **[T/F]** By including a lagged dependent variable, we implicitly include many lags of the explanatory variable.
11. (2.5 points) **[T/F]** Exogeneity essentially says that the disturbance must be independent of your explanatory variables.

12. (2.5 points) **[T/F]** Autocorrelation means two variables are strongly correlated.

13. (2.5 points) **[T/F]** Suppose the variable x_t is a random walk.
    True or false: $\Delta x_t = x_t - x_{t-1}$ is stationary.

14. (2.5 points) **[T/F]** To be a valid instrument, a variable must (1) correlate with the endogenous regressor and (2) correlate with the disturbance.

15. (2.5 points) **[T/F]** A *p*-value below 0.05 suggests the data do not support the null hypothesis.

16. (2.5 points) **[T/F]** In the model below, if the disturbance is autocorrelated, then OLS is biased and inconsistent for β_1.

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

17. (2.5 points) **[T/F]** In the first stage of two-stage least squares, we regress the outcome of interest on the instrumental variable.

18. (2.5 points) **[T/F]** The econometric model below allows the effect of income on health to depend upon the individual's age.

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

19. (2.5 points) **[T/F]** To be able to interpret a regression a causal, exogeneity must hold.

20. (2.5 points) **[T/F]** In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + u_t$, the parameter β_2 gives the effect of today's income on tomorrow's births.

Questions continue on the next page.

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.

21. (5 points) Explain how the terms *selection bias* and *omitted-variable bias* are related.

22. (5 points) What issues can non-stationary data cause? Briefly explain your answer.

23. (5 points) What is the purpose of an instrumental variable?

24. (5 points) We've all heard that *correlation is not causation*. Explain how correlation and causation are related.

25. (5 points) Where does the "least squares" part of the name in OLS come from?

26. (5 points) Should we be concerned about autocorrelation in cross-sectional data? Explain.

27. (5 points) In the following model, *Income* is measured in dollars, *Female* is an indicator, and *Age* is reported in years (since birth).

$$\text{Income}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i \times \text{Age}_i + u_i \tag{1}$$

Answer the following questions (you will have βs in your answers).

- What is the expected income for a 25-year old female?
- What is the average difference between a 50-year-old female and 50-year-old non-female?

28. (5 points) Suppose you wish to estimate the effect of health insurance (*Insurance*, a binary indicator below) on *Health*. Toward this goal, you plan to estimate the following regression using instrumental variables, where *Employed* (a binary indicator) is your instrument.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Insurance}_i + u_i \tag{2}$$

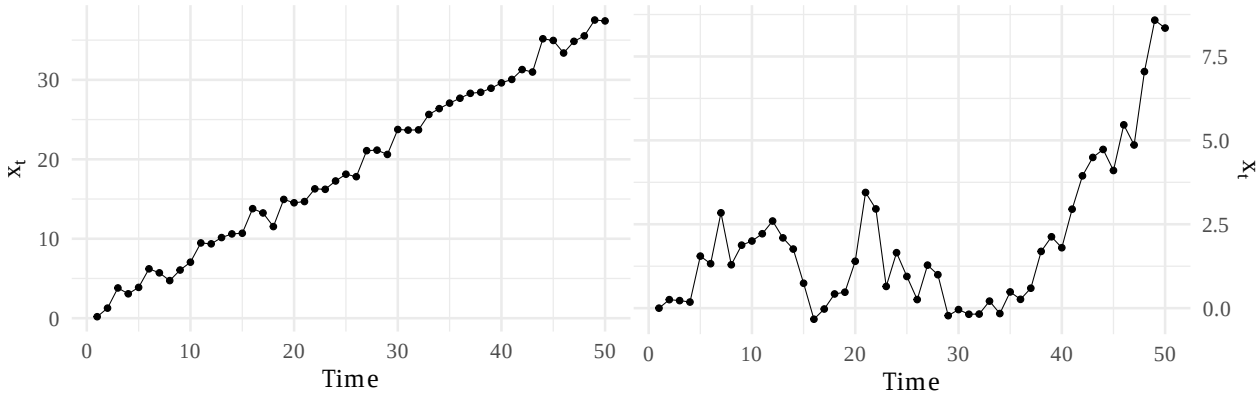
Does the *Employed* meet the requirements for a valid instrument? Explain your answer.

29. (5 points) Define and explain the fundamental problem of causal inference.

30. (5 points) For the model below, explain why we might be concerned about reverse causality.

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

3 Analyzing graphs (20 points)



(a) (b)

Figure 1: Stationarity

31. (10 points) In Figure 1 (above), each of the two subfigures depicts a 50-period time-series plot for a different random variable x_t.

    For *each* of the figures, discuss whether you believe the variable is (1) mean stationary and (2) variance stationary. Explain your reasoning.

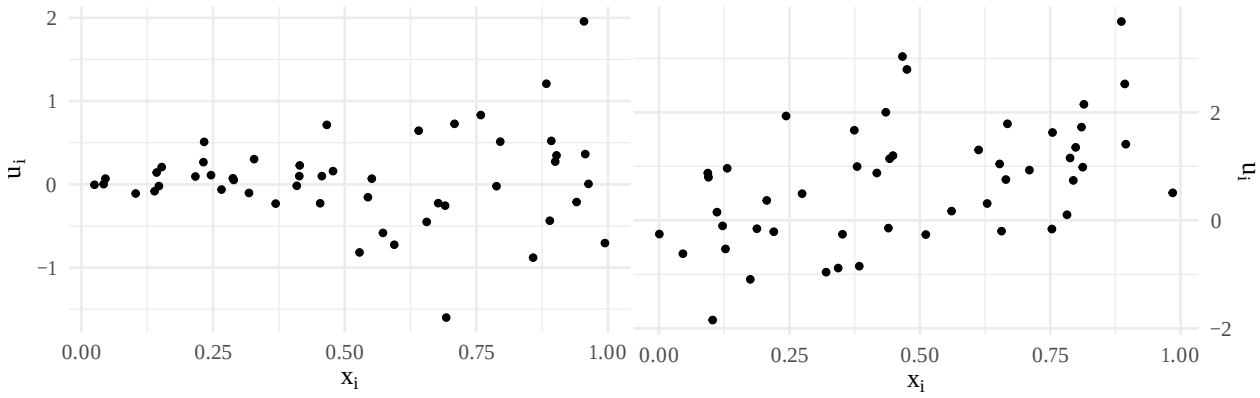


Figure 2: Assumptions

32. (10 points) In Figure 2 (above), each of the subfigures depicts the relationship between an explanatory variable (x_i) on the horizontal axis and the disturbance (u_i) on the vertical axis—for 50 observations. Think of each subfigure as a separate dataset.

    For *each* of the figures, discuss whether you believe the figure suggests the data are (1) exogenous and (2) homoskedastic. Explain your reasoning.