EC 421

Final

Spring 2019

Full Name ←

UO ID ← KEY

Total points: /108

No phones, calculators, or outside materials.

True/False

32.5 points

Note: You do not need to explain to your answers **in this section**.

01. **[T/F]** (**2.5pts**) In the presence of omitted-variable bias, ordinary least squares (OLS) is still consistent.

F

02. **[T/F]** (**2.5pts**) In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$, only the current period's income affects the current period's number of births.

T

03. **[T/F]** (**2.5pts**) In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + u_t$, only the current period's income affects the current period's number of births.

F

04. **[T/F]** (**2.5pts**) In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$, only the current period's income affects the current period's number of births.

F

05. **[T/F]** (**2.5pts**) If an estimator is biased, then it is not consisent.

F

06. **[T/F]** (**2.5pts**) As long as the mean of a variable does not change with time, the variable is stationary.

F

07. **[T/F]** (**2.5pts**) With a *p*-value of 0.049 and a significance level of 0.05, we reject the null hypothesis.

T

08. **[T/F]** (**2.5pts**) Random walk walks are stationary.

F

09. **[T/F]** (**2.5pts**) If the average *untreated* outcome ($\cancel{text}Y_{0i}$) in the treated group equals the average *untreated* outcome in the control group, then we do not need to worry about selection bias.

T

10. **[T/F]** (**2.5pts**) Exogeneity requires $E[u_i^2] = \sigma^2$.

F

11. **[T/F]** (**2.5pts**) Randomized experiments typically have issues with selection bias.

F

12. **[T/F]** (**2.5pts**) Causation has nothing to do with correlation.

F

13. **[T/F]** (**2.5pts**) Instrumental variables estimates can be consistent when OLS is inconsistent.

T

Short definitions

33 points

14. (**3pts**) Define the "standard error of an estimator".

The standard error of an estimator gives the standard deviation of the estimator's distribution.

15. (**3pts**) Define "autocorrelation".

Autocorrelation means a variable is correlated with itself — e.g., an individual's previous outcomes/shocks correlate with current outcomes/shocks.

16. (**3pts**) Define "nonstationarity".

Nonstationarity means a random variable is "stable" (behaves well) throughout time.

Alternatively: - Mean unaffected by time, $E[X_t] = E[X_{t+k}]\ \forall k$
- Variance unaffected by time, $Var(X_t) = Var(X_{t+k})\ \forall k$
- Covariance unaffected by time, $Cov(X_t, X_{t+k}) = Cov(X_s, X_{s+k})$ $\forall t, s, k$

17. (**3pts**) Define "p-value".

A p-value gives the probability of observing our outcome — or a more extreme outcome — under the null hypothesis.

18. (**3pts**) Define "causality".

Causality tells us the effect of an event by comparing the outcome of a universe w/. the event to the outcome in a universe w/.out the event (the counterfactual).

19. (**3pts**) Define "selection bias".

Selection bias is a type of bias that occurs when our control group and treatment group are not comparable.

20. (**3pts**) Define "the fundamental problem of causal inference"

The fundamental problem of causal inference is that we cannot simultaneously observe Y_{1i} (treated) and Y_{0i} (control) for the same individual.

21. (**3pts**) What is a "random walk"?

A random walk is a time-series process such that

$$X_t = X_{t-1} + \varepsilon_t$$

(Unnecessary: And ε_t is a stationary error term.)

22. (**6pts**) What are the two requirements for a *valid* instrument? Briefly define each requirement.

1. Exogenous: The instrument only affects Y_i (outcome) through X_i — not directly or through other variables.
2. Relevant: The instrument correlates w/. our variable of interest (X_i).

23. (**3pts**) What is "heteroskedasticity"?

Heteroskedasticity occurs when the variance of our disturbance is not constant, ie, $Var(\varepsilon_i) = \sigma_i^2$ and $\sigma_i^2 \neq \sigma_j^2$ for some i,j.

Also will accept: when variance of the disturbance correlates w/. an explanatory variable.

May use u_i.

Short answer

42.5 points

24. Suppose you want to estimate the causal effect of education on future earnings.

a. (**6pts**) Explain why regressing earnings on education is not going to give you the answer you want (*i.e.*, the causal effect of education on earnings).

Regressing earnings on education will likely suffer from omitted-variable bias (selection bias).

There are omitted variables that correlate w/. education AND affect earnings. This situation can lead to bias.

b. (**4pts**) Now imagine you know there is a scholarship program that randomly gives out scholarships. Individuals who receive these scholarships. Explain how you could use instrumental variables to estimate the causal effect of education on earnings.

We could use 2SLS (or IV) where scholarship winning is our instrument for education.

If they check this box, award 2 pts and move on to question 26.

☐ Check this box if you want to skip all of question 25 and instead receive **2pts**.

25. Each part of this question refers to the following R output which results from estimating

```
lm(income ~ education + experience + female, data = wage_df)
```

```
#> # A tibble: 4 x 5
#>   term        estimate std.error statistic  p.value
#>   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
#> 1 (Intercept)    3177.     2606.      1.22 2.26e- 1
#> 2 education       952.      109.      8.74 7.66e-14
#> 3 experience     2898.       89.0    32.6  1.10e-53
#> 4 female        -1761.     1570.     -1.12 2.65e- 1
```

a. (**2pts**) Write down the model that we've estimated (do not include the estimates, just the βs).

$$\text{income}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{experience}_i + \beta_3 \text{female}_i + u_i$$

Note: must include disturbance for full credit.

b. (**3pts**) Carefully interpret the coefficient on `education`.

Holding all else constant, an additional year of education increases income by 952.

c. (**3pts**) Conduct a hypothesis test for the coefficient on `education`. Describe each step and the explain your conclusion.

Hypotheses - $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$

p-value - p-value = $7.66 \times 10^{-14} < 0.05$ $\therefore$ Reject H_0

Conclusion - There is statistically significant evidence at the 5% level that education affects income.

d. (**3pts**) Carefully interpret the coefficient on `female` (an indicator variable for *female*).

Holding all else equal, women make 1761 less than men.

e. (**2.5pts**) **True/False** Because the p-value on `female` is greater than 0.05, we can conclude there is no gender-based difference in income.

False.

f. (**2pts**) If we remove `female` from the regression, do you expect the R^2 to increase or decrease? Explain your answer.

Decrease. Adding more variables increases R^2 mechanically.

g. (**3pts**) What assumptions/conditions must be satisfied for us to be able to interpret the coefficient on `education` as causal?

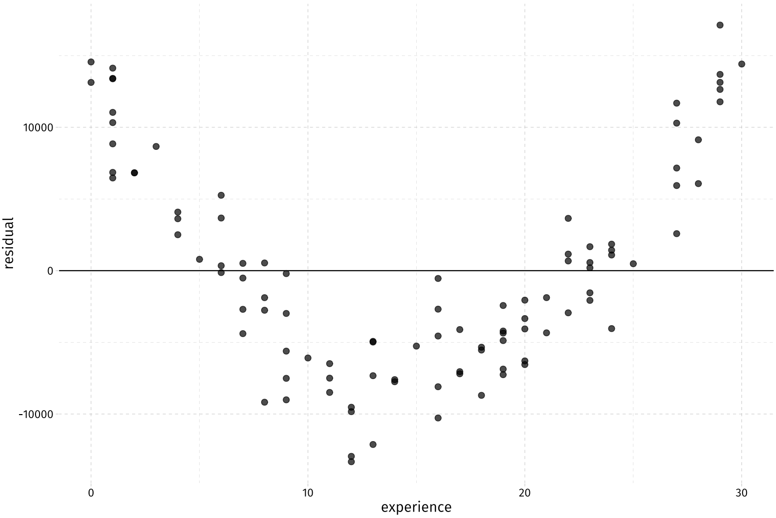
3 options:

Education must be exogenous

No selection bias

No omitted-variable bias

h. (**2pts**) Examine the following plot of the residuals from the previous regression (on the y axis) and `experience` (on the x axis). Describe any issues that suggests and how we might 'fix' them.



There is a trend in our residuals. This trend suggests misspecification.

☐ Check this box if you want to skip all of question 26 and instead receive **2pts**.

26. Consider the following model for the number of alcoholic drinks an individual consumes in a day.

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

a. (**2pts**) Carefully interpret the term β_1.

Holding all else constant, an additional drink yesterday increases the number of drinks today by β_1.

b. (**2pts**) Carefully interpret the term β_2.

Holding all else constant, an additional unit of income TODAY increases the number of drinks today by β_2.

c. (**2pts**) Carefully interpret the term β_3.

Holding all else constant, an additional unit of income YESTERDAY increases the number of drinks TODAY by β_3.

d. (**2pts**) What does $\beta_2 + \beta_3$ tell us?

The total effect of income on # of drinks.

e. (**2pts**) What does this model assume about the effect of income throughout time?

Income has no effect on drinking after t-1 (two days).

f. (**2pts**) What happens to our OLS-based estimate of β_1 if u_t is autocorrelated?

Biased and inconsistent