EC 421

Final Solutions

18 March 2019

Full Name ←

UO ID ←

No phones, calculators, or outside materials.

True/False

37.5 points

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

01. **[T/F]** (**2.5pts**) $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$ describes a static time series model.

FALSE

02. **[T/F]** (**2.5pts**) 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 **all** previous time periods on births in the current period.

FALSE

03. **[T/F]** (**2.5pts**) From the estimated model

$$\text{Births}_t = \hat{\beta}_0 + \hat{\beta}_1 \text{Income}_t + \hat{\beta}_2 \text{Income}_{t-1} + \hat{\beta}_3 \text{Income}_{t-2} + e_t$$

we can estimate the *total* effect of income on births as $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$.

FALSE

04. **[T/F]** (**2.5pts**) If the disturbances from two time periods t and s (with $t \neq s$) have non-zero covariance, *i.e.*, $\text{Cov}(u_t,\, u_s) \neq 0$, then we say that the disturbance is *heteroskedastic*.

FALSE

05. **[T/F]** (**2.5pts**) For dynamic models with lagged explanatory variables and autocorrelated disturbances, OLS we can trust OLS to be unbiased.

TRUE

06. **[T/F]** (**2.5pts**) If $\text{Var}(u_t) = 1$ for t in {1, ..., 10} and $\text{Var}(u_t) = 3$ for {11, ..., T}, then u_t is heteroskedastic.

TRUE

07. **[T/F]** (**2.5pts**) If $\text{Var}(u_t) = 1$ for t in {1, ..., 10} and $\text{Var}(u_t) = 3$ for {11, ..., T}, then u_t is nonstationary.

TRUE

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

FALSE

09. **[T/F]** (**2.5pts**) Selection bias refers to including observations with differing variances.

FALSE

10. [T/F] (**2.5pts**) *Prediction* focuses on estimating $\hat{\beta}$, while *casual inference* focuses on estimating $\hat{y}$.

FALSE

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

FALSE

12. [T/F] (**2.5pts**) Randomized experiment help us avoid selection bias by (approximately) balancing $Avg(y_{1,i}|D_i = 1)$ and $Avg(y_{1,i}|D_i = 0)$.

Too confusing—just grade for completion.

13. [T/F] (**2.5pts**) An instrumental variable is *exogenous* if it affects y (the outcome) through the endogenous x and through another explanatory variable w.

FALSE

14. [T/F] (**2.5pts**) The relevance requirement of instrumental variables is untestable.

FALSE

15. [T/F] (**2.5pts**) The exogeneity requirement of instrumental variables is untestable.

TRUE

Short Answer

62.5 points

Note: You will typically need to explain/justify your answers in this section.

16. (**2.5pts**) Write down an ADL(1,1) model for the effect of income on births.

Must include current income, lagged income, lagged births, and a disturbance:
$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_3 \text{Births}_{t-1} + u_t$

17. (**2.5pts**) Explain what negative autocorrelation in our disturbances means.
You can focus on an AR(1) process.

"Negative autocorrelation in our disturbances" means that our disturbances are negatively correlated in time. *E.g.*, in an AR(1) process, a large *positive* disturbance is likely to be followed by a large *negative* disturbance.

18. (**3pts**) In the following dynamic time-series model, u_t is first-order autocorrelated, *i.e.*,

$$\begin{aligned}\text{Health}_t &= \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Health}_{t-1} + u_t & \text{model, } t \\ \text{Health}_{t-1} &= \beta_0 + \beta_1 \text{Income}_{t-1} + \beta_2 \text{Health}_{t-2} + u_{t-1} & \text{model, } t-1 \\ u_t &= \rho u_{t-1} + \varepsilon_t & \text{AR}(1)\end{aligned}$$

where ε_t is "white noise"—independentaly and identically distributed with mean zero and variance σ^2_ε.

Explain why OLS will likely be biased for β_2—even if there are no omitted variables.

This model is likely to be biased—even without an omitted variables—because dynamic models with lagged outcome variables ***and*** autocorrelation violate contemporaneous exogeneity. The disturbance u_{t-1} affects Health_{t-1} (the model for t-1), and u_{t-1} also affects u_{t-1}. Thus Health_{t-1} (a regressor in model t) and u_t (the disturbance in model t) are likely correlated, which violates contemporaneous exogeneity (leading to potential bias).

19. (**3pts**) Suppose we are concerned about first-degree autocorrelation for the model

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

To test whether u_t has first-degree autocorrelation, we run a Breusch-Godfrey test and find an F statistic of 3.98, which has a *p*-value of approximately 0.0488. State the conclusion of this test and interpret the results.

With a *p*-value of 0.0488, we reject the null hypothesis (H_0 is *no autocorrelation of degree 1*). We conclude that there is statistically significant evidence, at the 5-percent level, or first-order autocorrelation.

20. (**2.5pts**) If we omit a variable that is autocorrelated, will our error term be autocorrelated? Explain your answer.

Yes. If we omit a variable, then that variable is in our error term. Thus, if a variable in the error term is autocorrelated, then the error term (containing that variable) is likely autocorrelated.

21. We have two individuals—Ali and Bob—and would like to use them to estimate the causal effect of college on income. Let $\text{Income}_{1,i}$ denote the income of individual i if she went to college and $\text{Income}_{0,i}$ if she did not.

$$\begin{aligned} \text{Income}_{1,\text{Ali}} &= 75,000 \qquad & \text{Income}_{0,\text{Ali}} &= 70,000 \\ \text{Income}_{1,\text{Bob}} &= 65,000 \qquad & \text{Income}_{0,\text{Bob}} &= 50,000 \end{aligned}$$

i. (**3pts**) Calculate the causal effect of education for Ali.

75,000 - 70,000 = 5,000

ii. (**3pts**) Calculate the causal effect of education for Bob.

65,000 - 50,000 = 15,000

iii. (**2.5pts**) Is the treatment effect constant? Briefly explain.

No. The two individuals have different treatment effects.

iv. (**3pts**) Bob went to college; Ali did not. Estimate the effect of college using the difference estimator—the difference in the mean of the treatment group (college) and the mean of the control group (no college).

65,000 - 70,000 = -5,000

v. (**4pts**) Do we have selection bias? Briefly explain your answer.

Yes. Our control group (Ali) does not provide a good *counterfactual* for the treatment group (Bob). Ali's non-college outcome (70,000) is much higher than Bob's non-college outcome (50,000).

22. (**2.5pts**) What is the fundamental problem of causal inference?

The fundamental problem of causal inference is that we cannot simultaneously observe outcomes from an individual as treatment and control.

Another way to say this: We cannot observe $y_{1,i}$ and $y_{0,i}$ at the same time.

23. (**4pts**) What problem does instrumental variables attempt to solve? How does it do it?

Instrumental variables attempts to solve violations of our requirement of exogeneity (can also say: an endogenous explanatory variable or omitted-variable bias).

To solve this problem, instrumental variables separates "good" (exogenous) variation in our endogenous variable from "bad" (endogenous) variation.

24. (**3pts**) What does it mean for an instrumental variable to be *relevant*?

An instrument is *relevant* if it *affects* our endogenous explanatory variable.

25. The probability limit of the instrumental variables estimator is

$$\text{plim } \hat{\beta}_1^{\text{IV}} = \beta_1 + \frac{\text{Cov}(z,\, u)}{\text{Cov}(z,\, x)}$$

where z is our instrument, u is the disturbance, and x is our endogenous variable.

i. (**4pts**) How do the two requirements of a valid instrument enter into this equation?

The denominator relates to **relevance**: If our instrument z is relevant for the endogenous variable x, then the denominator is not zero.

The numerator relates to **exogeneity**: If our instrument z is truly exogenous, then it is uncorrelated with the disturbance u, which makes the numerator zero.

ii. T/F (**2.5pts**) If we have a valid instrument, then $\hat{\beta}_1^{\text{IV}}$ is a consistent estimator for β_1.

TRUE

26. (**2.5pts**) Consider the random walk

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

where ε_t is stationary.

Take the difference between u_t and its lag. Is this difference stationary?

The difference: $u_t - u_{t-1} = u_{t-1} + \varepsilon_t - u_{t-1} = \varepsilon_t$, which is stationary (by assumption).

Therefore: yes, the difference is stationary.

27. Imagine we've estimated the model

$$\text{Income}_i = \beta_0 + \beta_1 (\text{Military service})_i + u_i$$

using two-stage least squares.

Income_i gives the income of individual i, and $\text{Military service}_i$ is a binary indicator variable for whether individual i served in the military.

Our instrument for military service is Male_i, an indicator variable for whether or not the individual is male.

The following two tables give the results of the first and second stage.

First-stage results (Outcome variable: Military service)

Term	Coef. estimate	Standard error	t stat.	p Value
Intercept	0.014	0.008	1.75	0.0801
Is Male	0.120	0.046	2.61	0.0091

Second-stage results (Outcome variable: Income)

Term	Coef. estimate	Standard error	t stat.	p Value
Intercept	31,153.71	7,829.2	3.98	<0.0001
Military service	1,104.85	329.10	3.36	0.0008

Questions on the next page...

i. (**3pts**) Write out the first-stage model that we estimated.

$$\text{Military Service}_i = \pi_0 + \pi_1 \text{Is Male}_i + v_i$$

Note: They can use any letter for the coefficients and the disturbance. It's also fine if they plugged in the estimates.

ii. (**3pts**) Interpret the first-stage results. (What do they say about gender differences in military service?)

The first-stage estimates suggest that there is a statically significant difference in military service by gender. Specifically, the intercept tells us approximately 1.4 percent of women serve in the military. The coefficient on *Is Male* tells us *an additional* 12 percent of men serve in the military (thus, 13.4 percent of men serve in the military). The difference between men and women is statistically significant.

iii. (**3pts**) Does it appear as though we have a *relevant* instrument? Explain your answer.

Yes. Our *t* statistic on *Is Male* in the first stage is strongly significant, which signals that our instrument is relevant for (strongly correlated with) our endogenous variable (military service).

iv. (**3pts**) Does it seem likely that our instrument is *exogenous*? Explain your answer.

No. For *Is Male* to be exogenous we need (1) *Is Male* to be uncorrelated with any other omitted variables that affect income (unlikely to be true: there are gender differences across many omitted and important variables, *e.g.*, education) and (2) *Is Male* can only affect income through military service (also unlikely—*e.g.*, labor-market discrimination).

Note: They only need to give a good reason *why* the instrument is not exogenous (they don't need this full answer).

v. (**3pts**) Assuming that we have a valid instrument, interpret the second-stage results.

Based upon our second-stage results, prior military service causes a statistically significant increase in income at the 5-percent level. This increase is approximately $1,104.85.

Extra credit: Venn diagrams(!)

- Each circle illustrates a variable.
- Overlap refers to the (share of) correlatation between two variables.
- Dotted borders denote *omitted* variables.

Figure A

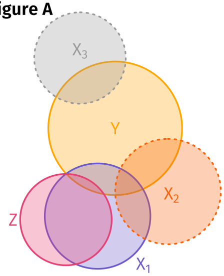


Figure B

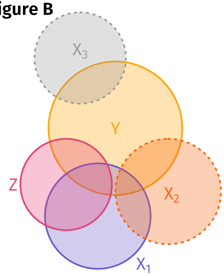


Figure C

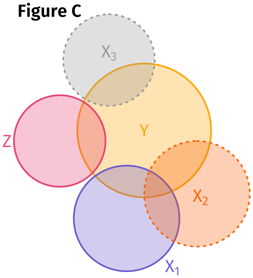
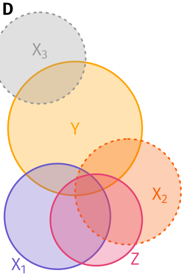


Figure D



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

EC$_1$. (**2pts**) Label the area(s) in **Figure A** that make us concerned about omitted-variable bias. **Label should denote the intersection between X_1 and X_2.**

EC$_2$. (**2pts**) In which figures is Z a valid instrument for X_1? **Only A**

EC$_3$. (**2pts**) In which figures is Z *relevant* for X_1? **A, B, and D**

EC$_4$. (**2pts**) In which figures is Z *exogenous* (with respect to X_1)? **Only A**

EC$_5$. (**2pts**) On the back of this page, draw a Venn diagram that has two valid instruments for an endogenous variable. Two instruments (*e.g.*, Z_1 and Z_2) intersect with X_1. They do not intersect any other variables. Z_1 and Z_2 can intersect with Y *only if* the intersection includes X_1.