Exam practice

https://pollev.com/jonathanold608

Midterm review

- Your questions
- Control variables: A 3D explanation.
- Logarithmic specifications
- Practice question
- Your questions

Your questions

Logarithmic specifications

Logarithmic specifications

y = a + b*x + e
y = a + b*log(x) + e
log(y) = a + b*x + e
log(y) = a + b*log(x) + e

TABLE 2.3 Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Practice questions

Practice question 1 (RCT, coefficient interpretation)

You are analyzing the results from a carefully constructed randomized evaluation of the effect of workplace incentives on worker productivity in a large moving company. You regress the log of worker output (measured in kg moved per hour) on a constant and an indicator for whether the individual was randomly selected to receive a financial incentive for high productivity.

A. **(4 marks)** The coefficient on the indicator variable is 0.20 with a standard error of 0.05. Explain, as if to a non-technical audience, what this estimate implies.

B. **(4 marks)** A colleague notices that because of the physical nature of the work, men are likely to be more productive at this job than women, all else equal. As such, your colleague claims, the worker's gender constitutes an omitted variable. You have data on each employee's gender so you rerun the regression including a gender dummy. How do you expect your estimate for the effect of incentives to change and why?

Practice question 1 (RCT, coefficient interpretation)

a. Individuals that were randomly selected to receive the financial incentives had a 0.2 higher log worker output. In other words, they had around 20% more output than individuals that were not selected. This difference is statistically significant at the 5% level.

b. Let us look at the OVB formula. Clearly, gender could be related to output, so $\partial>0$. If the incentive was really random, then gender is uncorrelated to the incentive, so $\pi=0$. Hence, $\partial\pi=0$, so there should be no effect of including gender as control.
Omitted variable bias is not a problem in RCTs.

Practice question 2 (OVB, coefficient interpretation)

We are interested to predict whether an individual was insured. We have data on their log wages and on their age. Interpret the coefficient on log wage in the first model. [**NOT EASY**]

	Model 1	Model 2
(Intercept)	1.84	-18.66 ***
	(1.25)	(0.68)
logwage	-0.18	2.57 ***
	(0.17)	(0.09)
age		-0.69 ***
		(0.02)
nobs	300	300
r.squared	0.00	0.86

Practice question 2 - Answer

a. In increase in wages by 1% is associated with an decrease in the probability of being insured by 0.18 percentage points. However, this difference is not significant at the 5% level.

Practice question 2 (OVB, coefficient interpretation)

b. Interpret the coefficient on age in column 2.
c. Using the OVB formula, find the coefficient in the auxiliary regression. If you don't find the coefficient, tell us what sign the coefficient has.

	Model 1	Model 2
(Intercept)	1.84	-18.66 ***
	(1.25)	(0.68)
logwage	-0.18	2.57 ***
	(0.17)	(0.09)
age		-0.69 ***
		(0.02)
nobs	300	300
r.squared	0.00	0.86

Practice question 2 - Answer

b. Individuals with one additional year of age are 69 percentage points less likely to be insured.

c. In the auxiliary regression, we regress the omitted variable on the included variable, so age on log wages.
Use the OVB formula:

$$\beta_S = \beta + \pi \text{ x } \delta \rightarrow \delta = (\beta_S - \beta)/\pi$$

In our case β_S=-0.18, π=-0.69 and β=2.57.

Therefore, δ=3.98.

The coefficient on log wages goes up, so we had downward bias. Therefore, $\pi \text{ x } \delta$ must be negative. Since π is negative, δ must be positive.

A systematic way to approach an exam question

1. **Think**: About the question, about the real world
2. Start with the numbers **you see**
   a. One-sentence summary
   b. Direction (positive or negative?)
   c. Statistical significance (significant or insignificant, at what level?)
   d. (Economic) magnitude (big or small?)
3. Then: Establish whether estimated relationship is **causal or not**
   a. What do the results **mean**? Correlation (interesting) or causality (policy-relevant)
   b. Is X-variable randomized? Do we have valid counterfactuals?
   c. If not: Do you expect **bias**? Of which sort (**OVB, reverse causality, bad controls**, …)?
   d. Find a plausible story around bias (using the OVB formula)

There are no traps!