API 202 Section

TF: Kelsey Pukelis

02/03/2023

Suppose policymakers are considering investing in an existing set of colleges to improve their quality. To help policymakers decide whether to invest large sums of money into this effort, you have been tasked with reading quantitative evidence on this issue. In short, you would like to know the answer to the following question:

Does the 'quality' of the college that students attend influence their subsequent earnings?

By college 'quality', we mean a more selective college. In general, selective colleges are considered higher quality because they have more (financial) resources to invest in students' education, higher 'quality' peers, etc.

Thankfully, you find an academic paper which addresses this question: Dale, S., & Krueger, A. (2002). "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *The Quarterly Journal of Economics*, 117(4), 1491-1527. https://doi.org/10.1162/003355302320935089

In this exercise, you will walk through how to read and interpret quantitative evidence from this paper based on what we have learned in the class so far about regression. At the end, we will come back to thinking about how we would communicate these results to policymakers.

---

1. Is the question we are interested in a causal question? [1-2 sentences]

Answer: Yes! We are interested in whether improving college quality will increase student's earnings later in life. We are interested in this causal question because we want to know whether or not it is worthwhile for the government to invest in college quality. If there is no causal relationship, then we would invest in school quality with no returns in terms of student earnings.

2. Write down a (short) bivariate PRF representing the main question we're interested in. (Use α's in your equation to remain consistent with the solutions.)

Answer:

$$earnings = \alpha_0 + \alpha_1 \cdot college.quality + u$$

Now, read this text from the paper's introduction:

"Past studies have found that students who attended colleges with higher average SAT scores or higher tuition tend to have higher earnings when they are observed in the labor market."

3. What are the two measures of college quality the authors refer to here?

Answer: (1) average SAT score of the college & (2) tuition of the college

Note: In the equation above, you could replace the very general, amorphous variable $college.quality$ with a particular <u>measure</u> of college quality like $avg.SAT.score$ of the college

The intro continues:

"Attending a college with a 100 point higher average SAT is associated with 3 to 7 percent higher earnings later in life (see, e.g., Kane [1998])."

4. How does this sentence relate to the regression equation we wrote down in Part 1? [1 sentence]

Answer: The Kane 1998 paper estimated the regression in a sample, and this sentence interprets the slope coefficient $\hat{\alpha}_1$ that the paper estimated.

The intro continues:

"As Kane notes, an obvious concern with this conclusion is that students who attend more elite colleges may have greater earnings capacity regardless of where they attend school. Indeed, the very attributes that lead admissions committees to select certain applicants for admission may also be rewarded in the labor market."

5. In your own words, restate the concern discussed in this paragraph. *[1-3 sentences]*

Answer: This paragraph is discussing the possibility of an omitted variable in the regression of earnings on college quality (here, referring to elite schools as an indicator of quality). In this context, the type of students who <u>attend elite colleges</u> are typically also the type of students who <u>earn more</u>, because of some omitted variable.

In general, if you can say something like, "the type of people who are more likely to ____ are also [more/less] likely to ____", then we may have omitted variable bias. This statement is describing a particular type of bias that economists are often concerned about called <u>selection bias</u>.

---

6. Going back to the statement: *"Attending a college with a 100 point higher average SAT is associated with 3 to 7 percent higher earnings later in life (see, e.g., Kane [1998])."* **Why are the authors using the language "is associated with" if we are interested in a causal question? [1-2 sentences]**

Answer: The authors do not make a causal statement because they are not convinced that the simple bivariate regression we wrote down in question 1 would estimate a causal parameter. Their main hesitation about making a causal claim is omitted variable bias, discussed above.

"Most past studies have used Ordinary Least Squares (OLS) regression analysis to attempt to control for differences in student attributes that are correlated with earnings and college qualities."

7. Give an example of a variable you might want to add to the short regression to account for omitted variable bias. Explain briefly why you choose this variable. [2-3 sentences]

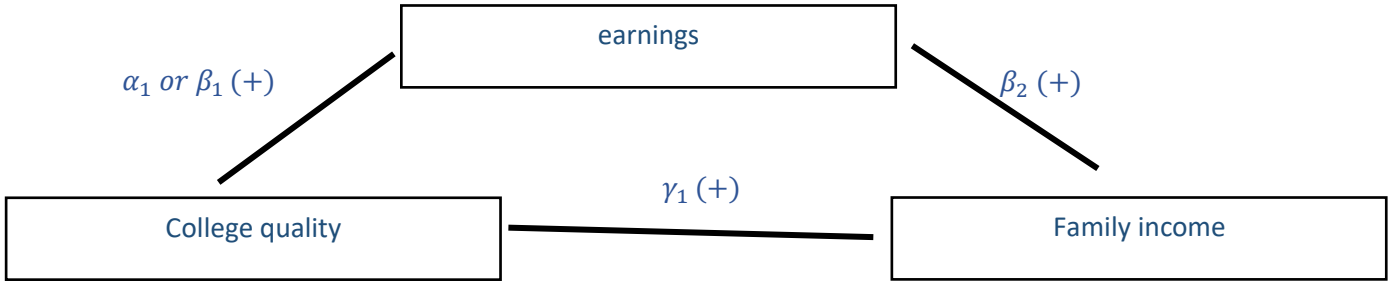
Answer: One example of an omitted variable is family income (family background). For example, children from higher income families are both more likely to go to a selective college (because of legacy admissions, standardized test prep, etc.) AND more likely to earn more as adults (through family connections, for example). In other words, family income is an example of a variable that is positively correlated with both earnings and college quality.

8. Write down the "long" PRF, including the variable you just chose. (Use β's in your equation.)

Answer:

$$earnings = \beta_0 + \beta_1 \cdot college.quality + \beta_2 \cdot family.income + v$$

9. Draw a triangular diagram relating the three variables. Label the arrows with the coefficients they correspond to and label the sign you expect them to have (+ or -).

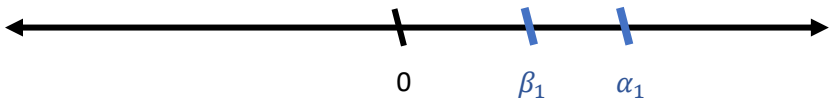


Where $college.quality = \gamma_0 + \gamma_1 \cdot family.income + w$

10. Sign the bias using the diagram below. Recall: $\boldsymbol{bias = \alpha_1 - \beta_1 = \gamma_1 * \beta_2}$

$\boldsymbol{bias = \alpha_1 - \beta_1 = \gamma_1 * \beta_2} = (+) * (+) = (+)$

Therefore: $\boldsymbol{\alpha_1 - \beta_1 > 0 \rightarrow \alpha_1 > \beta_1}$



11. Which estimate ($\widehat{\alpha}_1$or $\widehat{\beta}_1$) do you think would be closer to the true effect of college quality on earnings? Why? [1-2 sentences]

Answer: I expect the estimate $\widehat{\beta}_1$ ***to be closer to the true effect of college quality on earnings because the regression accounts for variation in earnings explained by parental background.*** $\widehat{\beta}_1$ ***will have the interpretation of the effect of college quality holding constant parental background.***

The introduction continues:

"But college admissions decisions are based in part on student characteristics that are unobserved by researchers and therefore not held constant in the estimated wage equations; if these unobserved characteristics are positively correlated with wages, then OLS estimates will overstate the payoff to attending a selective school."

12. What does this paragraph suggest about how we should interpret the coefficient on college selectivity from a super long equation which included many student characteristics? In other words, do the authors think that we can trust estimates from the super long equation? Why or why not? [2-3 sentences]

Answer: The authors think that, even after controlling for student characteristics, there are still variables omitted from the super long regression that would make the regression estimate of the effect of attending a selective school on earnings larger than the true causal effect. Because these variables are unobserved (i.e. not in any dataset), we may never be able to estimate the true causal effect of a selective school on earnings using this method.

Extra note: The paper goes on to develop a method that, they argue, estimates something closer to the true causal effect of school quality on earnings. Suppose there are two students who applied to the exact same set of colleges, were accepted and rejected by the same exact colleges, but, for some reason, one student went to College A (a very selective college) and the other went to College B (a not-so-selective college). The method then compares these two students' earnings to estimate the effect of attending a selective college. Intuitively, when we only make comparisons between very similar students, we get close to eliminating all possible omitted variables, thereby getting close to the causal effect of school quality on earnings.

13. Considering this entire exercise, how would you explain to policy makers the evidence on the effect of college quality on earnings? [2-3 sentences]

Answer: Although some studies find a positive association between school quality and earnings, this does not necessarily mean that improving school quality will increase students' future earnings. The positive relationship between school quality and earnings could be solely because of a third factor, like parental income; students from wealthier families are more likely to go to elite colleges and are also more likely to earn more as adults anyway. Before investing in improving school quality over other possible policies, we should look for evidence on school quality and earnings which accounts for these other potential factors.