

# Reading Assignment 7

Andrew Shaw

2025-10-12

## **1. What is the ingenious matching exercise in table 2.1? Why can we do this and forget about ALL possible omitted variables?**

By taking all observations in where there is uniformity across measurable characteristics so that each subgroup can be compared while looking only at the variable of interest's effect on the dependent outcome thereby reduce any influence from other characteristics on the outcome. Then subgroups can then be compared using weights depending on the size of those subgroups. This is a *ceteris paribus* situation so that comparisons are more like-with-like.

## **2. Why would it not have been prudent to compare students in Group A to those in Group D?**

The question being posed is whether a private school education has an effect on outcome later in life. The students in group D were all of the control group having only been accepted to public schools so there is no private school effect to try and isolate within group let alone across groups to group A.

## **3. Why does all of the weight of the regression to compare public to private rest on the effects in Group A and B and none of the weight come from comparisons within group C and group D?**

Again for similar reasons above, the two groups only offer information about either the control or the treatment group and thus there is no comparison to be made regarding benefits of private education.

## **4. Why are the weights different in regression than you'd think they might be if you look at simply the number of observations in the group?**

The addition of the group variable in the regression means the weight no looks within group at the variation between private and public schools not across group so the simple average is no longer sufficient for weighting.

## **5. What does transform the left-hand side of an equation by taking the natural log do to the betas on the right-hand side?**

This changes the interpretation of the coefficient to represent a percentage change in the dependent variable for a one unit change in the regressor.

**6. How can I use regression to get the price elasticity of demand explicitly from regression output?**

Transform both the dependent variable and the regressor into natural log format and then interpretation of the beta coefficient now becomes the elasticity.

**7. What can we learn from table 2.2? Explain the results in detail? Does private school matter as much as many assume it does?**

As more variables are included in the model and effects are isolated, the private school premium continues to decline while remaining statistically significant.

**8. Explain why we might be looking at a limited set here? What are the public schools included and how might they compare to other public schools?**

Because we're only looking at schools featured in the Barron's list of selectivity.

**9. Why might table 2.4 shed a little more light on the problem?**

Table shifts to looking at the campus-wide avg. SAT score of students which is continuous as opposed to a binary variable of private v. public. The binary variable captures a lot of information that isn't necessarily specific to private schools.

**10. We can't have all omitted variables? Why does sensitivity analysis give us greater confidence that we haven't omitted anything important?**

By comparing two models, one with a robust amount of controls and simpler model, the difference in the beta of the regressor of interest doesn't reveal a drastic difference there is unlikely a missing variable that could cause large fluctuations.

**11. How can we quantify the omitted variable bias differences when we include key controls compared to when we omit key controls?**

Take the  $\hat{\beta}_S - \hat{\beta}_L = \Delta OVB$ . There are other robustness checks which can be run as well, this being the most basic.

**12. Describe from table 2.5 what we can learn about omitted variable bias in this context?**

The table shows that earnings gap resulting from attending private school was actually bias coming from family background (parental income) and SAT scores of individuals.

**13. Why might we still have omitted variable bias despite our calculations?**

The calculations only describe how sensitive the coefficients are to hypothetical missing variables so there could be additional OVB not considered in the model.

**14. What is regression to the mean? How was it found as a phenomenon?**

It's a statistical property of correlated pairs of variables discovered when Galton studied heights of fathers and sons and compared to average height of population. He noticed that shorter than average parents tend to have slightly taller children and taller than average parents have slightly shorter children.