

## Section Week 9

The datafile Birthweight\_Smoking is from the 1989 linked National Natality-Mortality Detail files, which contains a census of infant births and deaths. The data in birthweight\_smoking.wf1 are for births in Pennsylvania in 1989. These data were provided by Professor Douglas Almond, Kenneth Chay, and David Lee and are a subset of the data used in their paper “The Costs of Low Birth Weight”, Quarterly Journal of Economics, August 2005, 120(3): 1031-1083. This exercise was adapted from Empirical Exercise 6.1 in Stock and Watson.

The file contains 3,000 observations on the variables described below

	Variable	Description
Birthweight and Smoking		
1	birthweight	birth weight of infant (in grams)
2	smoker	indicator equal to one if the mother smoked during pregnancy and zero, otherwise.
Mothers Attributes		
3	age	age
4	educ	years of educational attainment (more than 16 years coded as 17)
5	unmarried	indicator =1 if mother is unmarried
This Pregnancy		
6	alcohol	indicator=1 if mother drank alcohol during pregnancy
7	drinks	number of drinks per week
8	tripre1	indicator=1 if 1st prenatal care visit in 1st trimester
9	tripre2	indicator=1 if 1st prenatal care visit in 2nd trimester
10	tripre3	indicator=1 if 1st prenatal care visit in 2nd trimester
11	tripre0	indicator=1 if no prenatal visits
12	nprevist	total number of prenatal visits

1. Regress Birthweight on a constant and educ. Interpret the estimated effect of a year of education on birth weight.

**Answer:** When we compare people whose education differs by one year, the average value of birth weight is 28.75 grams larger among those whose education is one year higher. In other words, a one year increase in education is associated with a 28.75 gram heavier baby on average.

```
ls birthweight c educ
```

Dependent Variable: BIRTHWEIGHT  
Method: Least Squares  
Date: 05/25/18 Time: 14:40  
Sample: 1 3000  
Included observations: 3000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3011.814	64.96328	46.36179	0.0000
EDUC	28.75339	4.963751	5.792674	0.0000
R-squared	0.011069	Mean dependent var		3382.934
Adjusted R-squared	0.010739	S.D. dependent var		592.1629
S.E. of regression	588.9748	Akaike info criterion		15.59531
Sum squared resid	1.04E+09	Schwarz criterion		15.59931
Log likelihood	-23390.97	Hannan-Quinn criter.		15.59675
F-statistic	33.55507	Durbin-Watson stat		1.973108
Prob(F-statistic)	0.000000			

2. Does the coefficient of educ measure the causal effect of a year of education on birth weight?

**Answer:** No, because there are omitted variables in the regression. Smoking is an omitted variable in the regression because it is correlated with both birth weight and education.

3. Does the omission of the smoking variable lead to the coefficient on education to be positively or negatively biased? Assume education and smoking are negatively correlated and smoking and birth weight are negatively correlated.

**Answer:** We think of a model with no intercepts to make calculation easier (difference in means model). We ran the regression

$$bw = \gamma educ + \epsilon,$$

but the true data generating process is

$$bw = \beta educ + \delta smoker + \varepsilon.$$

We can then derive the direction of the omitted variable bias:

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}] &= \mathbb{E} \left[ \frac{\sum educ * bw}{\sum educ^2} \right] \\
&= \mathbb{E} \left[ \frac{\sum educ * (\beta educ + \delta smoker + \varepsilon)}{\sum educ^2} \right] \\
&= \mathbb{E} \left[ \beta + \delta \frac{\sum educ * smoker}{\sum educ^2} + \frac{\sum educ * \varepsilon}{\sum educ^2} \right] \\
&= \beta + \delta \mathbb{E} \left[ \frac{\sum educ * smoker}{\sum educ^2} \right] + 0. \\
&\approx \beta + \underbrace{\delta}_{(-)} \underbrace{\frac{Cov(educ, smoker)}{Var(educ)}}_{(-)}
\end{aligned}$$

$\beta$ , the relationship between education and bw, is positive.  $\delta$ , the relationship between smoker and bw, is negative.  $Cov(educ, smoker)$  is negative. Therefore the bias is positive and the estimated coefficient is larger than the actual parameter. Once we add the omitted variable smoker we would expect the coefficient on educ to decrease.

- Now, regress Birthweight on a constant, educ, and smoker. Interpret the estimated effect of a year of education on birth weight.

Answer: When we compare people whose education differs by one year but have the same value for smoker, the average value of birth weight is 19.02 grams larger among those whose education is one year higher. In other words, a one year increase in education is associated with a 28.75 gram heavier baby on average holding smoker constant.

```
ls birthweight c educ smoker
```

Dependent Variable: BIRTHWEIGHT				
Method: Least Squares				
Date: 05/25/18 Time: 14:40				
Sample: 1 3000				
Included observations: 3000				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3181.736	67.44085	47.17817	0.0000
EDUC	19.02989	5.047227	3.770366	0.0002
SMOKER	-228.9741	27.65098	-8.280868	0.0000
R-squared	0.033190	Mean dependent var		3382.934
Adjusted R-squared	0.032544	S.D. dependent var		592.1629
S.E. of regression	582.4474	Akaike info criterion		15.57335
Sum squared resid	1.02E+09	Schwarz criterion		15.57936
Log likelihood	-23357.03	Hannan-Quinn criter.		15.57551
F-statistic	51.44207	Durbin-Watson stat		1.966260
Prob(F-statistic)	0.000000			

- You ask them if the coefficient of the treatment measures the causal effect of the treatment on the outcome.

Answer: No, because there are still omitted variables in the regression, such as alcohol and total number of prenatal visits.