Research question: causal effect of mother's smoking during pregnancy on birth weight of a child.

Data is from the 1989 linked National Natality-Mortality Detail files, which contains a census of infant births and deaths. The data set contains all births in Pennsylvania in 1989.
These data were provided by Porfessor 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.
The file contains 3,000 observations on the variables described below

Variable Description

Birthweight and Smoking	
`birthweight`	birth weight of infant (in grams)
`smoker`	indicator equal to one if the mother smoked during pregnancy and zero, otherwise.
Mother's Attributes	
`age`	age
`educ`	years of educational attainment (more than 16 years coded as 17)
`unmarried`	indicator =1 if mother is unmarried
This Pregnancy	
`alcohol`	indicator=1 if mother drank alcohol during pregnancy
`tripre1`	indicator=1 if 1st prenatal care visit in 1st trimester
`tripre2`	indicator=1 if 1st prenatal care visit in 2nd trimester
`tripre3`	indicator=1 if 1st prenatal care visit in 3rd trimester
`tripre0`	indicator=1 if no prenatal visits
`nprevist`	total number of prenatal visits

```
summ birthweight nprevist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 birthweight |      3000    3382.934    592.1629         425       5755
    nprevist |      3000    10.99167    3.672069           0         35

regress birthweight smoker, robust

Linear regression                               Number of obs =    3000
                                                F(  1,  2998) =   89.21
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0286
                                                Root MSE      =  583.73

------------------------------------------------------------------------------
             |               Robust
 birthweight |   Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-253.2284  26.81039    -9.45   0.000     -305.797   -200.6597
       _cons |  3432.06   11.89053   288.64   0.000     3408.746    3455.374
------------------------------------------------------------------------------
```

$$\underset{(11.89053)}{\textit{Birthweight} = \textit{3432.06}} \; - \; \underset{(26.81039)}{\textit{253.2284 smoker}}$$

What could cause omitted variable bias?

- *Alcohol consumption during pregnancy?*
- *Poor attention to health?*

How would you argue for this?

```
corr smoker alcohol nprevist
(obs=3000)

             |   smoker  alcohol nprevist
-------------+---------------------------
      smoker |   1.0000
     alcohol |   0.1209   1.0000
    nprevist |  -0.1086  -0.0425   1.0000
```

```
regress birthweight smoker alcohol, robust

Linear regression                               Number of obs =    3000
                                                F(  2,  2997) =   44.75
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0288
                                                Root MSE      =  583.77

------------------------------------------------------------------------------
             |               Robust
 birthweight |   Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-250.8034   26.86935    -9.33   0.000    -303.4877   -198.1192
     alcohol |-57.60132   77.38742    -0.74   0.457    -209.339    94.13651
       _cons |3432.703    11.93957   287.51   0.000     3409.29    3456.114
------------------------------------------------------------------------------
```

Birthweight = 3432.70 -250.8 smoker - 57.6 alcohol
(11.93) (26.87) (77.39)

Birthweight = 3432.06 -253.2 smoker
(11.89) (26.81)

```
regress birthweight smoker nprevist, robust

Linear regression                               Number of obs =    3000
                                                F(  2,  2997) =   89.16
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0728
                                                Root MSE      =  570.39

------------------------------------------------------------------------------
             |               Robust
 birthweight |   Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-218.8294   25.98922    -8.42   0.000    -269.7879   -167.8709
    nprevist |34.10394    3.608863     9.45   0.000     27.02784    41.18004
       _cons |3050.527    43.69276    69.82   0.000     2964.857    3136.198
------------------------------------------------------------------------------
```

Birthweight = 3050.5 -218.8 smoker + 34.1 nprevisit
(43.69) (25.99) (3.6)

Birthweight = 3432.06 -253.2 smoker
(11.89) (26.81)

What regression do you prefer? Why?

```
regress birthweight smoker nprevist tripre0 tripre1 tripre2 tripre3, r
note: tripre3 omitted because of collinearity

Linear regression                               Number of obs =    3000
                                                F(  5,  2994) =   37.19
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0767
                                                Root MSE      =  569.48

------------------------------------------------------------------------------
             |               Robust
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-219.4964   25.7869    -8.51   0.000    -270.0582   -168.9346
    nprevist |  33.8179   4.292605     7.88   0.000     25.40115    42.23465
     tripre0 |-416.6767   163.0092    -2.56   0.011     -736.298   -97.05544
     tripre1 |-115.7889   73.78332    -1.57   0.117      -260.46    28.88227
     tripre2 |-105.4244   74.30345    -1.42   0.156    -251.1154    40.26654
     tripre3 |          0  (omitted)
       _cons |3167.192   70.54545    44.90   0.000     3028.869    3305.514
------------------------------------------------------------------------------
```

What happens? Why?

```
regress birthweight smoker nprevist  tripre1 tripre2 tripre3 , robust

Linear regression                               Number of obs =    3000
                                                F(  5,  2994) =   37.19
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0767
                                                Root MSE      =  569.48

------------------------------------------------------------------------------
             |               Robust
 birthweight |   Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-219.4964   25.7869    -8.51   0.000    -270.0582   -168.9346
    nprevist | 33.8179   4.292605     7.88   0.000     25.40115    42.23465
     tripre1 |300.8879   156.1035     1.93   0.054    -5.193062    606.9688
     tripre2 |311.2523   154.6024     2.01   0.044     8.114665    614.3899
     tripre3 |416.6767   163.0092     2.56   0.011     97.05544     736.298
       _cons |2750.515   146.7299    18.75   0.000     2462.814    3038.217
------------------------------------------------------------------------------
```

```
test tripre1 tripre2 tripre3

 ( 1)  tripre1 = 0
 ( 2)  tripre2 = 0
 ( 3)  tripre3 = 0

       F(  3,  2994) =    2.41
            Prob > F =    0.0649
-------------------------

test tripre1 tripre2 tripre3 nprevist

 ( 1)  tripre1 = 0
 ( 2)  tripre2 = 0
 ( 3)  tripre3 = 0
 ( 4)  nprevist = 0

       F(  4,  2994) =   23.90
            Prob > F =    0.0000
```

```
regress birthweight smoker nprevist unmarried, robust

Linear regression                               Number of obs =    3000
                                                F(  3,  2996) =   74.75
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0886
                                                Root MSE      =  565.61

------------------------------------------------------------------------------
             |               Robust
 birthweight |  Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |-176.2122  26.70666    -6.60   0.000    -228.5775   -123.847
    nprevist |29.62307    3.58212     8.27   0.000      22.5994   36.64673
   unmarried |-187.2587  27.66867    -6.77   0.000    -241.5102  -133.0071
       _cons |3133.957   44.11278    71.04   0.000     3047.46    3220.452
------------------------------------------------------------------------------

. corr smoker unmarried
(obs=3000)

             |   smoker unmarr~d
-------------+------------------
      smoker |   1.0000
   unmarried |   0.2377   1.0000
```

Results of Regressions of Babies' weight at birth on Mothers' smoking, other actions and characteristics

Dependent variable: birthweight

Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Smoking	-253.2** (26.8)	-250.8** (26.9)	-218.8** (26.0)	-219.5** (25.8)	-176.2** (26.7)
Alcohol		-57.6 (77.4)			
Number of pre-natal visits			34.1** (3.6)	33.8** (4.3)	29.6** (3.6)
Dummies for Time of first visit	no	no	no	yes	no
Not married					–187.3** (27.7)
Summary statistics					
Root MSE	583.7	583.7	570.4	569.4	565.6
$\bar{R}^2$	0.028	0.028	0.073	0.077	0.089
n	3000	3000	3000	3000	3000
F-statistic testing coefficients on dummies for first visit (*p*-value)				2.41 (0.65)	
F-statistic testing coefficients on dummies for first visit and number of visits (*p*-value)				23.9** (0.00)	

The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test. Robust standard errors are given in parentheses under coefficients. All regressions include intercept

Return to education example:

```
regress wage educ female, robust

Linear regression                               Number of obs =     526
                                                F(  2,   523) =   69.10
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2588
                                                Root MSE      =  3.1855

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5064521   .0598956     8.46   0.000     .3887867    .6241176
      female |  -2.273362   .2702033    -8.41   0.000    -2.804179   -1.742545
       _cons |   .6228168   .7286843     0.85   0.393    -.8086909    2.054324
------------------------------------------------------------------------------
```

Wage=0.62+0.506 Educ-2.27 Female

```
regress wage educ female, robust

Linear regression                               Number of obs =     526
                                                F(  2,   523) =   69.10
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2588
                                                Root MSE      =  3.1855

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5064521   .0598956     8.46   0.000     .3887867    .6241176
      female |  -2.273362   .2702033    -8.41   0.000    -2.804179   -1.742545
       _cons |   .6228168   .7286843     0.85   0.393    -.8086909    2.054324
------------------------------------------------------------------------------

regress wage educ, robust

Linear regression                               Number of obs =     526
                                                F(  1,   524) =   78.09
                                                Prob > F      =  0.0000
                                                R-squared     =  0.1648
                                                Root MSE      =  3.3784

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5413593   .0612596     8.84   0.000     .4210146    .6617039
       _cons |  -.9048516   .7254795    -1.25   0.213    -2.330057    .5203539
------------------------------------------------------------------------------
```

```
correlate educ female
(obs=526)

             |     educ   female
-------------+------------------
        educ |   1.0000
      female |  -0.0850   1.0000
```

- *Can omitting Female cause an Omitted Variable Bias? Why?*
- *What other factors can cause OVB?*
- *Can Experience cause OVB?*

```
correlate educ exper
(obs=526)

             |     educ    exper
-------------+------------------
        educ |   1.0000
       exper |  -0.2995   1.0000
```

```
regress wage educ exper, robust

Linear regression                               Number of obs =     526
                                                F(  2,   523) =   50.32
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2252
                                                Root MSE      =   3.257

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6442721   .0651869     9.88   0.000     .5162117    .7723324
       exper |   .0700954   .0109943     6.38   0.000      .048497    .0916938
       _cons |  -3.390539   .8648747    -3.92   0.000    -5.089595   -1.691484
------------------------------------------------------------------------------
```

Results of Regressions of Hourly Earnings on Years of Education and Worker Characteristics

Dependent variable: Hourly earnings

Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Years of education	.54** (.061)	.51** (0.06)	0.64** (0.065)	0.59** (0.061)	0.57** (0.057)
Female		-2.27** (0.27)			-1.81** (0.25)
Experience			0.07** (0.010)	0.02* (0.010)	0.025* (0.009)
Tenure				0.169** (0.029)	0.141** (0.027)
Intercept	-.90 (.72)	0.62 (0.72)	-3.39 (0.86)	-2.87 (0.80)	–1.57 (0.83)
Summary statistics					
Root MSE	3.37	3.19	3.26	3.08	2.96
$\bar{R}^2$	0.165	0.25	0.23	0.31	0.36
n	526	526	526	526	526
F-statistic testing coefficients on Tenure and Experience 0 (*p*-value)					26.86 (0.000)

The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test. Standard errors are given in parentheses under coefficients.

```
reg wage educ female exper tenure,robust

  Linear regression                               Number of obs =     526
                                                  F(  4,   521) =   44.57
                                                  Prob > F      =  0.0000
                                                  R-squared     =  0.3635
                                                  Root MSE      =  2.9576

  ------------------------------------------------------------------------------
               |               Robust
          wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
          educ |    .5715048   .0612172     9.34   0.000     .4512418    .6917677
        female |   -1.810852   .254156     -7.12   0.000    -2.310149   -1.311556
         exper |    .0253959   .0098057     2.59   0.010     .0061323    .0446594
        tenure |    .1410051   .027955      5.04   0.000     .0860867    .1959234
         _cons |   -1.567939   .8259341    -1.90   0.058    -3.190509    .0546317
  ------------------------------------------------------------------------------
  test exper=tenure=0
   ( 1)  exper - tenure = 0
   ( 2)  exper = 0

         F(  2,   521) =   26.86
              Prob > F =    0.0000
```