

ECON4150 - Introductory Econometrics Seminar 3

Stock and Watson AEE6.2, AEE7.3

April 28, 2015

```
clear all
cd M:\pc\Desktop\courses\introductory_econometrics\seminar_3
use "CollegeDistance.dta"
cap log close
log using AEE6_2AEE7_3.log,replace
set more off
```

Variables in the set

```
ed          Years of Education Completed (See below)
female      1 = Female/0 = Male
black       1 = Black/0 = Not-Black
Hispanic    1 = Hispanic/0 = Not-Hispanic
bytest      Base Year Composite Test Score.
dadcoll     1 = Father is a College Graduate/ 0 = Father is not a College Graduate
momcoll     1 = Mother is a College Graduate/ 0 = Mother is not a College Graduate
incomehi    1 = Family Income > $25,000 per year/ 0 = Income ? $25,000 per year.
ownhome     1= Family Owns Home / 0 = Family Does not Own Home
urban       1 = School in Urban Area / = School not in Urban Area
cue80       County Unempolyment rate in 1980
stwmfg80    State Hourly Wage in Manufacturing in 1980
dist        Distance from 4yr College in 10's of miles
tuition     Avg. State 4yr College Tuition in $1000's
```

AEE6.2, a)

```
//describe the data
describe
//summary statistics
summ
```

```
//a
```

```
reg ed dist, r
```

Linear regression

```
Number of obs =    3796
F( 1, 3794) =    29.83
Prob > F      =    0.0000
R-squared     =    0.0074
Root MSE     =    1.8074
```

```
-----+-----
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
dist	-.0733727	.0134334	-5.46	0.000	-.0997101	-.0470353
_cons	13.95586	.0378112	369.09	0.000	13.88172	14.02999

```
-----+-----
```

```
// the estimated slope is -.0733727
```

```
//we will need the estimated coefficients later on, so we already store them
```

```
estimate store rega, title(univariate regression)
```

b)

```
reg ed dist bytest female black hispanic incomehi ownhome dadcoll cue80 stwmfg80, robust
```

```
Linear regression
```

```
Number of obs = 3796
F( 10, 3785) = 197.68
Prob > F      = 0.0000
R-squared     = 0.2788
Root MSE     = 1.5425
```

```
-----
```

	ed	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
dist		-.0315387	.0116616	-2.70	0.007	-.0544023 -.0086752
bytest		.0938201	.0029804	31.48	0.000	.0879768 .0996634
female		.145408	.0503939	2.89	0.004	.0466061 .2442098
black		.367971	.0675359	5.45	0.000	.2355608 .5003812
hispanic		.3985196	.0738763	5.39	0.000	.2536785 .5433608
incomehi		.3951984	.0619207	6.38	0.000	.2737972 .5165996
ownhome		.1521313	.0649193	2.34	0.019	.0248511 .2794115
dadcoll		.6961324	.0707602	9.84	0.000	.5574006 .8348641
cue80		.0232052	.00931	2.49	0.013	.0049521 .0414583
stwmfg80		-.0517777	.0196751	-2.63	0.009	-.0903526 -.0132029
_cons		8.827518	.2413001	36.58	0.000	8.354427 9.300609

```
-----
```

```
// the estimated slope is -.0315387
```

```
estimate store regb, title(multivariate Regression)
```

c)

coefficient in the multivariate regression is around half the coefficient in the univariate case. It is reasonable to believe that the regression in a) suffers from an omitted variable bias. The second regression attenuates this bias, adding regressors that are probably correlated both with *dist* and *ed*.

d)

```
/*  
  first you can install the external package estout.  it is easy as you only write:  
  "ssc install estout" and it will install automatically.  esttab or estout produce  
  nice tables of stored results  
*/
```

```
ssc install estout
```

```
esttab rega regb, r2 ar2
```

```
-----  
                rega          regb  
                b            b  
-----  
dist            -.0733727    -.0315387  
bytest                   .0938201  
female                   .145408  
black                    .367971  
hispanic                  .3985196  
incomehi                  .3951984  
ownhome                   .1521313  
dadcoll                   .6961324  
cue80                     .0232052  
stwmfg80                 -.0517777  
_cons                 13.95586    8.827518  
-----  
r2                      .0074496    .2788377  
r2_a                     .007188    .2769323  
-----
```

d)

- Regression a: $R^2 = 0.0074496$, $R_{adj}^2 = 0.007188$
- Regression b: $R^2 = 0.2788377$, $R_{adj}^2 = 0.2769323$

As $R_{adj}^2 = 1 - \frac{n-1}{n-k-1} \frac{SSR}{SST}$ and $R^2 = 1 - \frac{SSR}{SST}$, for large n (in our case $n=3796$)

$$\frac{n-1}{n-k-1} \approx 1 \text{ and } R_{adj}^2 \approx R^2$$

e), f)

text question

g)

```
/* Bob:
Black = 1
dist = 2
bytest = 58
incomehi = 1 as fam income is 26.000>25000
ownhome= 1
cue80 = 7.5
stwmfg80 = 9.75
*/

// you can pick up the stored coefficients directly from the regression:

display _b[_cons]+_b[dist]*2+_b[bytest]*58+_b[incomehi]*1+_b[black] ///
+_b[ownhome]*1+_b[cue80]*7.5+_b[stwmfg80]*9.75

14.790515
```

h)

```
display _b[_cons]+_b[dist]*4+_b[bytest]*58+_b[incomehi]*1+_b[black] ///  
+_b[ownhome]*1+_b[cue80]*7.5+_b[stwmfg80]*9.75
```

14.727437

reg ed dist

Source	SS	df	MS	
Model	93.0256754	1	93.0256754	Number of obs = 3796
Residual	12394.3568	3794	3.266831	F(1, 3794) = 28.48
				Prob > F = 0.0000
				R-squared = 0.0074
				Adj R-squared = 0.0072
Total	12487.3825	3795	3.29048287	Root MSE = 1.8074

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ed						
dist	-.0733727	.0137498	-5.34	0.000	-.1003304	-.046415
_cons	13.95586	.0377241	369.95	0.000	13.88189	14.02982

```
display _b[dist]+1.96*_se[dist]
-.04642309
```

```
display _b[dist]-1.96*_se[dist]
-.10032232
```

/*

variable dist is Distance from 4yr College in 10's of miles.
 Increase of 0.15 year when distance decreases by 20 miles implies a decrease of
 $0.15/2 = 0.075$ year when distance increases by 10 miles,
 as the value -0.075 is included in the 95% confident interval, the advocacy groups' claim
 is consistent with the estimated regress.

b)

```
qui reg ed dist bytest female black hispanic incomehi ownhome dadcoll cue80 ///  
stwmfg80 urban tuition, robust  
est sto reg3, title(Regression 3)  
/*qui is short for the command quietly and it means that stata runs the  
regression but without showing the results. Thus it looks like nothing happens  
in stata, however, isomg estto afterwards the results are picked up and stored.  
*/
```

b)

esttab rega regb reg3, se r2 ar2

	(1)	(2)	(3)
	ed	ed	ed
dist	-0.0734*** (0.0134)	-0.0315** (0.0117)	-0.0336** (0.0127)
bytest		0.0938*** (0.00298)	0.0945*** (0.00299)
female		0.145** (0.0504)	0.146** (0.0504)
black		0.368*** (0.0675)	0.353*** (0.0690)
hispanic		0.399*** (0.0739)	0.345*** (0.0776)
incomehi		0.395*** (0.0619)	0.403*** (0.0620)
ownhome		0.152* (0.0649)	0.149* (0.0653)
dadcoll		0.696*** (0.0708)	0.700*** (0.0709)
cue80		0.0232* (0.00931)	0.0272** (0.00957)
stwmfg80		-0.0518** (0.0197)	-0.0440* (0.0200)
urban			0.0621 (0.0636)
tuition			-0.189 (0.0991)
_cons	13.96*** (0.0378)	8.828*** (0.241)	8.862*** (0.244)
N	3796	3796	3796
R-sq	0.007	0.279	0.280
adj. R-sq	0.007	0.277	0.277

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

b)

```
/*  
  As noted in the previous exercises from the first to the second specification  
  the change in coefficient is large. However, different base specifications,  
  do not change much the coefficient.  
*/  
  
// Answering again to point a) with the first base specification reg2  
  
display -0.0315+1.96*0.0117  
-.008568  
display -0.0315-1.96*0.0117  
-.054432  
  
/*  
  [-0.054,-0.008] Thus the groups claim of the coefficient being -0.075 is no  
  longer in the confidence interval. Thus the claim is not consistent with the  
  estimated regression. the same conclusion would be reach for the third specification.  
*/
```

c)

//c) yes, in both specifications the coefficients for black are positive and statistically significant

log close