ECON4150 - Introductory Econometrics Seminar 5

Stock and Watson add.E8.2; add.E9.2

April 28, 2015

```
clear all
//cd M:\pc\Desktop\courses\introductory_econometrics\seminar_5
cd /Users/andreapapini/Desktop/Courses/introducory_econometrics/seminar_5/
use "CollegeDistance.dta"
cap log close
log using AEE8_2_AEE9_2.log , replace
set more off
```

Name	Desrciption
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. (These are achievement tests given to high school seniors in the sample)
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

```
//summary statistics
summ
```

Variable	Obs	Mean	Std. Dev.	Min	Max
female	3796	.5453109	.4980083	0	1
black	3796	.1925711	.394371	0	1
hispanic	3796	.1498946	.3570151	0	1
bytest	3796	51.00193	8.819251	28.95	71.36
dadcoll	3796	.2020548	.4015858	0	1
momcoll	3796	.1393572	.3463645	0	1
ownhome	3796	.8192835	.3848338	0	1
urban	3796	.243941	.4295141	0	1
cue80	3796	7.654874	2.86577	1.4	24.9
stwmfg80	3796	9.556499	1.364411	6.59	12.15
dist	3796	1.724921	2.133836	0	16
tuition	3796	.9131396	.2835778	.43418	1.40416
incomehi	3796	.2863541	.4521164	0	1
ed	3796	13.82929	1.813969	12	18

a)

```
 reg ed tuition dist bytest incomehi ownhome dadcoll ///
momcoll cue80 stwmfg80 black hispanic female in 1/3796 , r

Linear regression                               Number of obs =    3796
                                                F( 12,  3783) =  168.48
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2836
                                                Root MSE      =  1.5378

------------------------------------------------------------------------------
             |               Robust
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.1910519   .0985259    -1.94   0.053    -.3842209    .0021171
        dist |  -.0366613   .0120749    -3.04   0.002    -.0603352   -.0129874
      bytest |   .0930377    .003014    30.87   0.000     .0871284    .0989469
    incomehi |   .3718305   .0622177     5.98   0.000     .2498471    .4938138
     ownhome |   .1385475   .0649795     2.13   0.033     .0111492    .2659459
     dadcoll |   .5709712   .0763028     7.48   0.000     .4213726    .7205698
     momcoll |   .3778102   .0834999     4.52   0.000      .214101    .5415193
       cue80 |   .0286753   .0095229     3.01   0.003     .0100049    .0473458
    stwmfg80 |  -.0425003   .0199355    -2.13   0.033    -.0815857   -.0034148
       black |   .3506095   .0674301     5.20   0.000     .2184066    .4828125
    hispanic |   .3617649   .0764184     4.73   0.000     .2119397    .5115902
      female |   .1429742   .0502718     2.84   0.004     .0444118    .2415366
       _cons |   8.920823   .2434585    36.64   0.000       8.4435    9.398145
------------------------------------------------------------------------------
predict yhata
est sto rega
```

a

```
/*
 if distance increases from 2 to 3 or 6 to 7 education
 is expected to be reduced by  0.03666 years in both cases
*/
```

b

```
gen lned = ln(ed)
 reg lned tuition dist bytest incomehi ownhome dadcoll ///
 momcoll cue80 stwmfg black hispanic female in 1/3796 , r

Linear regression                               Number of obs =    3796
                                                F( 12,  3783) =  173.89
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2853
------------------------------------------------------------------------------
        lned |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.0139382   .0070081    -1.99   0.047    -.0276783   -.0001982
        dist |  -.0026072   .0008651    -3.01   0.003    -.0043032   -.0009111
      bytest |   .0066561   .0002133    31.21   0.000     .0062379    .0070742
    incomehi |   .0265197      .0044     6.03   0.000     .0178931    .0351463
     ownhome |   .0098332   .0046395     2.12   0.034      .000737    .0189295
     dadcoll |   .0405374   .0053518     7.57   0.000     .0300446    .0510302
     momcoll |   .0266016   .0058414     4.55   0.000     .0151491    .0380541
       cue80 |   .0020357   .0006768     3.01   0.003     .0007088    .0033626
    stwmfg80 |  -.0028642   .0014142    -2.03   0.043    -.0056368   -.0000916
       black |   .0261676   .0048091     5.44   0.000     .0167389    .0355963
    hispanic |   .0259986   .0054098     4.81   0.000     .0153922    .0366049
      female |   .0103059   .0035664     2.89   0.004     .0033137    .0172981
       _cons |   2.265819   .0172772   131.15   0.000     2.231946    2.299693
------------------------------------------------------------------------------
predict yhatb
est sto regb
```

b

```
/*
  if distance increases from 2 to 3 or 6 to 7 education
  is expected to be reduced by  100%*0.0026=0.26%.
 */
```

```
gen dist2=dist^2
 reg ed tuition dist dist2 bytest incomehi ownhome dadcoll ///
> momcoll cue80 stwmfg black hispanic female in 1/3796
      Source |       SS       df       MS              Number of obs =    3796
-------------+------------------------------           F( 13,  3782) =  115.61
       Model |  3551.12556    13  273.163505           Prob > F      =  0.0000
    Residual |  8936.25695  3782  2.36283896           R-squared     =  0.2844
-------------+------------------------------           Adj R-squared =  0.2819
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.1928193   .1008433    -1.91   0.056     -.390532    .0048933
        dist |  -.0811732   .0255486    -3.18   0.001    -.1312635   -.0310829
       dist2 |   .0046413   .0023109     2.01   0.045     .0001106     .009172
      bytest |   .0926367   .0031866    29.07   0.000      .086389    .0988843
    incomehi |   .3694975   .0607082     6.09   0.000     .2504734    .4885216
     ownhome |    .14327    .066738     2.15   0.032      .012424     .274116
     dadcoll |   .5611581   .0738274     7.60   0.000     .4164127    .7059035
     momcoll |   .3777022   .0814925     4.63   0.000     .2179287    .5374757
       cue80 |   .0259537   .0099544     2.61   0.009     .0064371    .0454702
    stwmfg80 |  -.0425539   .0202001    -2.11   0.035    -.0821581   -.0029497
       black |   .3339309   .0716843     4.66   0.000     .1933873    .4744745
    hispanic |   .3333104   .0785278     4.24   0.000     .1793494    .4872713
      female |   .1433144   .0504168     2.84   0.004     .0444676    .2421611
       _cons |   9.012167   .2555647    35.26   0.000     8.511109    9.513225
------------------------------------------------------------------------------
 predict yhatc
 est sto regc
```

```
/*
   if distance increases from 2 to 3  education
   is expected to be reduced by  (-.0812 * 3 + .0046 * 3^2)- (-.0812 * 2 + .0046 * 2^2)=-0.058
   if distance increases from 6 to 7  education
  is expected to be reduced by  (-.0812 * 7 + .0046 * 7^2)- (-.0812 * 6 + .0046 * 6^2)=-0.0212
*/
```

d

```
 /*
>    the two regressions differ just with respect an additional regressor (dist2) add in
>    the specification in c. Then, we prefer regression in c to the regression in
>    a as the coefficient dist2 is statistically significant different from 0 at
>    a 5 percent level (p-value=0.045)
> */
.
```

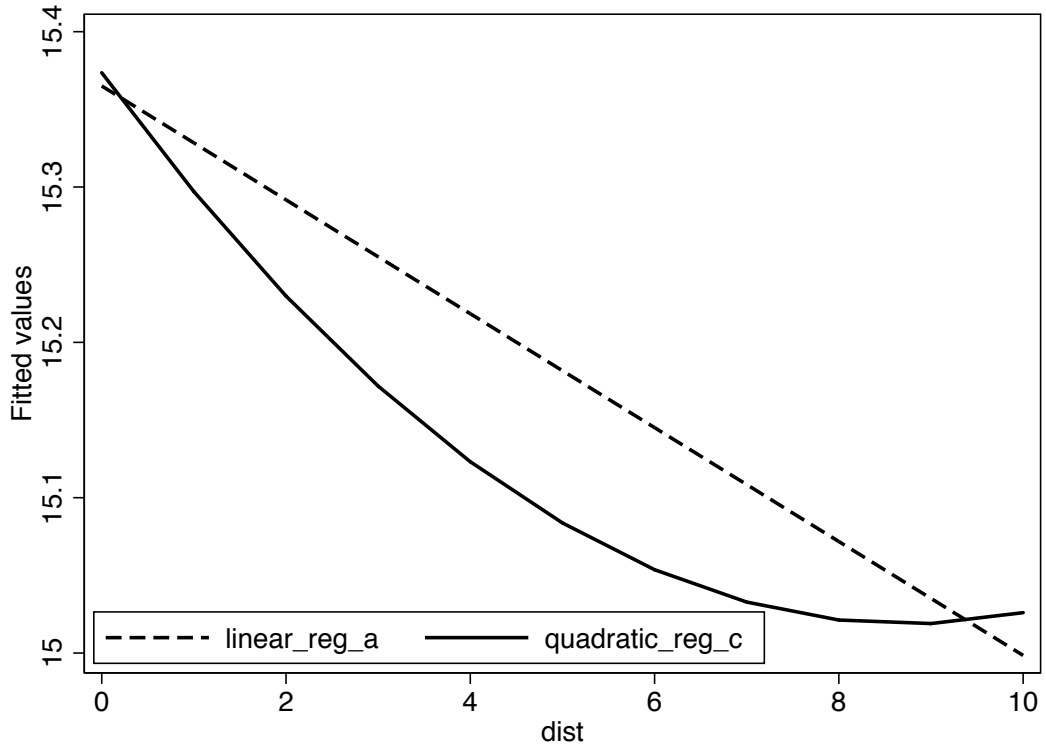
e

```
//i
 qui regress ed dist female bytest tuition black hispanic incomehi ownhome dadcoll ///
                      momcoll cue80 stwmfg80, robust
 matrix b=e(b)
 matrix list b
 gen y_predict_a=b[1,1]*dist+b[1,2]*1+b[1,3]*58+b[1,4]*0.95+b[1,5]*0+b[1,6]*1+ ///
                  b[1,7]*0+b[1,8]*0+b[1,9]*1+b[1,10]*1+b[1,11]*7.1+b[1,12]*10.06+b[1,13]

 qui regress ed dist dist2 female bytest tuition black hispanic incomehi ownhome ///
                       dadcoll momcoll cue80 stwmfg80, robust

  matrix b=e(b)
 matrix list b
  gen y_predict_c=b[1,1]*dist+b[1,2]*dist2+b[1,3]*1+b[1,4]*58+b[1,5]*0.95+b[1,6]*0+ ///
 b[1,7]*1+b[1,8]*0+b[1,9]*0+b[1,10]*1+b[1,11]*1+b[1,12]*7.1+b[1,13]*10.06+b[1,14]
sort dist
two (line  y_predict_a dist if dist<=10 , lwidth(medthick) lpattern(solid) lcolor(black)) ///
(line  y_predict_c dist if dist<=10  , lwidth(medthick) lpattern(dash) lcolor(black)) ///
, scheme(s1color) legend(pos(7) ring(0) label(1 "linear_reg_a") label(2 "quadratic_reg_c"))
```

e




```
//i
/*
  it would not change, the lines would just be shifted,
  as there is no interaction term between female and distance,
  then the estimated relation of distance on education does not depend on the gender
 */
```

```
//ii
summ if dist>10

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |        44    .4772727    .5052578          0          1
       black |        44           0           0          0          0
    hispanic |        44    .5227273    .5052578          0          1
      bytest |        44    49.87727    7.562153      37.16      66.73

....

/*
  The relation switches sign around the level of dist=10 (the minimum is reached at
  0.0811732/(2*0.0046413)=8.7446621). it is not plausible.
  just 44 obs, slightly more than 1%. Hence it is highly imprecise
*/
```

```
gen dadxmomcoll = dadcoll*momcoll
reg ed tuition dist dist2 female bytest incomehi ownhome dadcoll ///
 momcoll cue80 stwmfg black hispanic dadxmomcoll, r
Linear regression                               Number of obs =    3796
                                                F( 14,  3781) =  145.73
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2854
                                                Root MSE      =  1.5363

------------------------------------------------------------------------------
             |               Robust
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.1939714   .0985584    -1.97   0.049    -.3872042   -.0007387
        dist |  -.0810001    .025094    -3.23   0.001    -.1301992   -.0318011
       dist2 |   .0046773   .0020564     2.27   0.023     .0006455    .0087091
      female |   .1406184   .0502133     2.80   0.005     .0421707    .2390661
      bytest |   .0925664   .0030234    30.62   0.000     .0866388    .0984939
    incomehi |   .3623156   .0622537     5.82   0.000     .2402615    .4843697
     ownhome |   .1412131   .0649487     2.17   0.030     .0138752    .2685511
     dadcoll |   .6538031    .087084     7.51   0.000      .483067    .8245392
     momcoll |   .5693549   .1218052     4.67   0.000     .3305445    .8081652
       cue80 |   .0257697     .00959     2.69   0.007     .0069677    .0445716
    stwmfg80 |  -.0415432   .0199035    -2.09   0.037    -.0805658   -.0025206
       black |   .3305619   .0683148     4.84   0.000     .1966244    .4644994
    hispanic |   .3297465   .0779131     4.23   0.000     .1769907    .4825024
 dadxmomcoll |  -.3664802   .1639813    -2.23   0.025    -.6879805   -.0449799
       _cons |    9.00197   .2500197    36.01   0.000     8.511783    9.492157
------------------------------------------------------------------------------
 est sto reg4
```

f

```
/*
  the interactive effect of having both parents that completed college. It is estimated
  to be negative.
  next question makes the significant clear
*/
```

```
/*
   lincom computes point estimates, standard errors, t or z statistics, p-values, and
   confidence  intervals for linear combinations of coefficients
   after any estimation command.
 */
display "Jane vs Mary"
Jane vs Mary
 lincom _b[dadcoll]
 ( 1)  dadcoll = 0
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .6538031    .087084     7.51   0.000      .483067    .8245392
------------------------------------------------------------------------------
display "Alexis vs Mary"
Alexis vs Mary
 lincom _b[momcoll]
 ( 1)  momcoll = 0
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .5693549   .1218052     4.67   0.000     .3305445    .8081652
------------------------------------------------------------------------------
```

```
 display "Bonnie vs Mary"
Bonnie vs Mary
lincom _b[dadcoll]+_b[momcoll]+_b[dadxmomcoll]
 ( 1)  dadcoll + momcoll + dadxmomcoll = 0
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .8566778   .0951835     9.00   0.000     .6700618    1.043294
------------------------------------------------------------------------------
/*
  the interaction between momcoll and dadcoll allow the partial joint effect of having
  both parents that finish college to be smaller or larger than the sum of the effect.
  In this case it is smaller, .8566778 against, .5693549: estimated effect of only momcoll
  .6538031, estimated effect of only dadcoll
*/
```

```
gen incomedist = incomehi*dist
gen incomedist2 = incomehi*dist2
reg ed tuition dist dist2 female bytest incomehi incomedist incomedist2 ownhome dadcoll ///
momcoll cue80 stwmfg black hispanic dadxmomcoll
      Source |       SS       df       MS              Number of obs =    3796
-------------+------------------------------           F( 16,  3779) =   94.73
       Model |  3574.58278    16  223.411424            Prob > F      =  0.0000
    Residual |  8912.79973  3779  2.35850747            R-squared     =  0.2863

       Total |  12487.3825  3795  3.29048287            Root MSE      =  1.5357
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.2099784   .1010178    -2.08   0.038     -.408033   -.0119238
        dist |  -.1095309   .0287061    -3.82   0.000    -.1658119     -.05325
       dist2 |   .0064744   .0024925     2.60   0.009     .0015876    .0113613
      female |    .141463   .0503871     2.81   0.005     .0426744    .2402516
      bytest |   .0927566   .0031851    29.12   0.000      .086512    .0990012
    incomehi |   .2172968   .0906892     2.40   0.017     .0394922    .3951013
  incomedist |   .1244186   .0641478     1.94   0.053    -.0013491    .2501863
 incomedist2 |   -.008659   .0071381    -1.21   0.225    -.0226538    .0053359
     ownhome |   .1437389    .066693     2.16   0.031     .0129811    .2744966
     dadcoll |   .6627368   .0844088     7.85   0.000     .4972456    .8282281
     momcoll |   .5674681   .1172585     4.84   0.000     .3375721    .7973641
       cue80 |   .0260482   .0099476     2.62   0.009     .0065451    .0455514
    stwmfg80 |  -.0419249   .0201874    -2.08   0.038    -.0815042   -.0023456
       black |    .333128   .0716511     4.65   0.000     .1926494    .4736065
    hispanic |   .3230637   .0785375     4.11   0.000     .1690838    .4770436
 dadxmomcoll |  -.3556964    .161458    -2.20   0.028    -.6722495   -.0391432
       _cons |   9.042179   .2560325    35.32   0.000     8.540204    9.544154
------------------------------------------------------------------------------

est sto reg5
```

```
. reg ed tuition dist dist2 female bytest incomehi ownhome dadcoll ///
 momcoll cue80 stwmfg black hispanic dadxmomcoll

      Source |       SS       df       MS              Number of obs =    3796
-------------+------------------------------           F( 14,  3781) =  107.84
       Model |  3563.30281    14  254.521629            Prob > F      =  0.0000
    Residual |   8924.0797  3781  2.36024324            R-squared     =  0.2854
-------------+------------------------------           Adj R-squared =  0.2827
       Total |  12487.3825  3795  3.29048287            Root MSE      =  1.5363
------------------------------------------------------------------------------
          ed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tuition |  -.1939714   .1007892    -1.92   0.054    -.3915779     .003635
        dist |  -.0810001   .0255346    -3.17   0.002    -.1310632   -.0309371
       dist2 |   .0046773   .0023097     2.03   0.043     .0001489    .0092056
      female |   .1406184   .0504031     2.79   0.005     .0417986    .2394383
      bytest |   .0925664    .003185    29.06   0.000     .0863219    .0988109
    incomehi |   .3623156   .0607572     5.96   0.000     .2431955    .4814357
     ownhome |   .1412131   .0667075     2.12   0.034      .010427    .2719993
     dadcoll |   .6538031   .0843096     7.75   0.000     .4885064    .8190998
     momcoll |   .5693549   .1172733     4.85   0.000     .3394298    .7992799
       cue80 |   .0257697   .0099493     2.59   0.010     .0062631    .0452762
    stwmfg80 |  -.0415432   .0201939    -2.06   0.040    -.0811353   -.0019512
       black |   .3305619   .0716603     4.61   0.000     .1900654    .4710584
    hispanic |   .3297465   .0785003     4.20   0.000     .1758394    .4836537
 dadxmomcoll |  -.3664802   .1613445    -2.27   0.023     -.682811   -.0501494
       _cons |    9.00197   .2554637    35.24   0.000      8.50111     9.50283
------------------------------------------------------------------------------
```

h

```
//  I calculate the homosckedasticity-only F statistic
display (( 8924.0797- 8912.79973)/2)/( 8912.79973 /3779)
// |2.39|>2.36 so jointly significant at a 10% level, but not at a 5% level as Fc=3.00
```

```
est tab reg*, se
```

Variable	rega	regb	regc	reg4	reg5
tuition	-.1910519	-.01393823	-.19281935	-.19397145	-.20997838
	.09852587	.00700812	.10084335	.09855838	.10101775
dist	-.03666128	-.00260716	-.08117324	-.08100015	-.10953093
	.01207487	.00086508	.02554857	.02509396	.02870611
bytest	.09303769	.00665608	.09263667	.09256637	.09275657
	.00301402	.00021327	.00318662	.00302335	.00318506
incomehi	.37183046	.02651971	.36949749	.36231564	.21729675
	.06221765	.00440001	.06070825	.06225371	.09068921
ownhome	.13854754	.00983323	.14327003	.14121312	.14373886
	.06497954	.00463954	.06673802	.06494874	.06669299
dadcoll	.57097117	.04053742	.56115815	.65380312	.66273682
	.07630279	.00535185	.07382742	.08708396	.08440881
momcoll	.37781017	.0266016	.37770219	.56935486	.56746807
	.08349989	.00584136	.0814925	.12180525	.11725846
cue80	.02867534	.00203567	.02595365	.02576966	.02604824
	.00952285	.00067678	.00995444	.00958996	.00994758
stwmfg80	-.04250025	-.00286418	-.04255391	-.04154322	-.0419249
	.01993554	.00141416	.02020013	.01990348	.02018744
black	.35060952	.02616763	.33393089	.3305619	.33312796
	.06743014	.00480911	.07168429	.06831485	.0716511
hispanic	.36176494	.02599856	.33331038	.32974654	.32306372
	.07641841	.00540978	.0785278	.07791315	.07853747
female	.14297422	.0103059	.14331438	.14061844	.14146301
	.05027176	.00356639	.0504168	.05021326	.05038715
dist2			.0046413	.00467727	.0064744
			.0023109	.00205641	.00249254
dadxmomcoll				-.3664802	-.35569638
				.16398125	.16145795
incomedist					.12441861
					.06414783
incomedist2					-.00865897
					.00713806
_cons	8.9208225	2.2658194	9.012167	9.00197	9.0421792

add.EE9.3

```
use CollegeDistancewest.dta,replace
gen dist2 = dist^2
gen dadxmomcoll = dadcoll*momcoll
gen incomedist = incomehi*dist
gen incomedist2 = incomehi*dist2
```

```
reg ed tuition dist dist2 female bytest income* ownhome dadcoll ///
momcoll cue80 stwmfg black hispanic dadxmomcoll
```

Source	SS	df	MS		
Model	617.124199	16	38.5702625	Number of obs	= 943
Residual	2052.52585	926	2.2165506	F(16, 926)	= 17.40
Total	2669.65005	942	2.83402341	Prob > F	= 0.0000
				R-squared	= 0.2312
				Adj R-squared	= 0.2179
				Root MSE	= 1.4888

ed	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tuition	-.5226686	.2605872	-2.01	0.045	-1.034079	-.0112586
dist	-.0916798	.046738	-1.96	0.050	-.1834044	.0000449
dist2	.0040874	.0033019	1.24	0.216	-.0023927	.0105674
female	.0505693	.0990857	0.51	0.610	-.1438892	.2450278
bytest	.0732997	.0065034	11.27	0.000	.0605365	.0860629
incomehi	.4070321	.1704902	2.39	0.017	.0724401	.7416241
incomedist	.0045501	.0944787	0.05	0.962	-.180867	.1899673
incomedist2	-.0000224	.0065375	-0.00	0.997	-.0128524	.0128075
ownhome	.1992296	.1300655	1.53	0.126	-.0560278	.4544869
dadcoll	.4412696	.1477206	2.99	0.003	.1513635	.7311756
momcoll	.283049	.2373124	1.19	0.233	-.1826835	.7487815
cue80	.0452626	.0249629	1.81	0.070	-.0037278	.0942531
stwmfg80	.0307996	.0498222	0.62	0.537	-.066978	.1285771
black	.0671427	.2211574	0.30	0.762	-.3668852	.5011706
hispanic	.1955382	.1169679	1.67	0.095	-.0340148	.4250911
dadxmomcoll	.1422522	.3159081	0.45	0.653	-.4777267	.7622311
_cons	9.227512	.5618685	16.42	0.000	8.124829	10.3302

```
est tab reg5 reg6, se
-------------------------------------
    Variable |    reg5         reg6
-------------+-----------------------
     tuition | -.20997838    -.5226686
             |  .10101775    .26058723
        dist | -.10953093    -.0916798
             |  .02870611    .04673798
       dist2 |   .0064744    .00408737
             |  .00249254    .00330188
      female |  .14146301    .05056933
             |  .05038715    .09908567
      bytest |  .09275657    .07329969
             |  .00318506    .00650343
    incomehi |  .21729675    .40703214
             |  .09068921     .1704902
  incomedist |  .12441861    .00455015
             |  .06414783    .09447866
 incomedist2 | -.00865897   -.00002242
             |  .00713806    .00653746
     ownhome |  .14373886    .19922956
             |  .06669299    .13006553
     dadcoll |  .66273682    .44126957
             |  .08440881    .14772063
     momcoll |  .56746807    .28304897
             |  .11725846    .23731241
       cue80 |  .02604824    .04526264
             |  .00994758     .0249629
    stwmfg80 |  -.0419249    .03079956
             |  .02018744    .04982223
       black |  .33312796    .06714272
             |   .0716511    .22115743
    hispanic |  .32306372    .19553817
             |  .07853747    .11696791
 dadxmomcoll | -.35569638    .14225222
             |  .16145795    .31590812
       _cons |  9.0421792    9.2275123
```