Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\k19056473\Downloads\fees.smcl
  log type:  smcl
 opened on:  19 Nov 2019, 19:04:12

1 . use "C:\Users\k19056473\Downloads\TuitionFees.dta"

2 .
3 . *1. Estimation of the DD model
4 .
5 . gen logApp=ln(applications)

6 . gen logPop=ln(Poptotal)

7 . gen afterinc = (year>=2012) * (app_country == "England")

8 . gen Eng = app_country=="England"

9 . gen male = gender=="Male"

10. tabulate age_band, generate (age_dummy)
```

age_band	Freq.	Percent	Cum.
18 and under	**27,627**	**27.99**	**27.99**
19	**24,580**	**24.91**	**52.90**
20	**21,607**	**21.89**	**74.79**
21 and over	**24,877**	**25.21**	**100.00**
Total	**98,691**	**100.00**	

```
11. summarize logApp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
logApp	**98,691**	**3.884657**	**1.488375**	**1.609438**	**8.809863**

```
12. xi: reg logApp Eng i.year afterinc male age_dummy2 age_dummy3 age_dummy4 logPop, vce
  > (cluster app_country)
  i.year            _Iyear_2008-2015    (naturally coded; _Iyear_2008 omitted)

  Linear regression                               Number of obs     =     98,691
                                                  F(0, 1)           =          .
                                                  Prob > F          =          .
                                                  R-squared         =     0.1908
                                                  Root MSE          =      1.339

                                (Std. Err. adjusted for 2 clusters in app_country)
```

logApp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Eng	**-.1331743**	**3.259322**	**-0.04**	**0.974**	**-41.54679**	**41.28044**
_Iyear_2009	**.0676512**	**.0190722**	**3.55**	**0.175**	**-.1746847**	**.3099871**
_Iyear_2010	**.1646741**	**.0243298**	**6.77**	**0.093**	**-.1444649**	**.4738132**
_Iyear_2011	**.1921612**	**.0384266**	**5.00**	**0.126**	**-.2960954**	**.6804179**
_Iyear_2012	**.2948299**	**.0129002**	**22.85**	**0.028**	**.130917**	**.4587428**
_Iyear_2013	**.3156208**	**.0056578**	**55.79**	**0.011**	**.2437318**	**.3875098**
_Iyear_2014	**.2482423**	**.0064197**	**38.67**	**0.016**	**.166672**	**.3298126**
_Iyear_2015	**.253904**	**.0059056**	**42.99**	**0.015**	**.1788658**	**.3289422**
afterinc	**-.2065389**	**.0017811**	**-115.96**	**0.005**	**-.2291693**	**-.1839084**
male	**-.0732275**	**.0138894**	**-5.27**	**0.119**	**-.2497092**	**.1032542**
age_dummy2	**-.32385**	**1.004761**	**-0.32**	**0.802**	**-13.09055**	**12.44285**
age_dummy3	**-1.177184**	**1.14271**	**-1.03**	**0.491**	**-15.69669**	**13.34232**
age_dummy4	**-2.874516**	**4.564605**	**-0.63**	**0.642**	**-60.87333**	**55.1243**
logPop	**.5466306**	**1.408375**	**0.39**	**0.764**	**-17.34847**	**18.44173**
_cons	**1.340122**	**5.994739**	**0.22**	**0.860**	**-74.83026**	**77.5105**

```
13.
14. *2. Graph
15.
16. collapse (sum) applications, by (year app_country)

17. gen treat = app_country=="England"

18. gen logApp=ln(applications)

19. twoway (line logApp year if treat == 0) (line logApp year if treat == 1)

20.
21. *3. country-specific  linear trend
22.
23. clear

24. use "C:\Users\k19056473\Downloads\TuitionFees.dta"

25. gen logApp=ln(applications)

26. gen logPop=ln(Poptotal)

27. gen afterinc = (year>=2012) * (app_country == "England")

28. gen Eng = app_country=="England"

29. gen male = gender=="Male"

30. gen CounTrend=Eng*year

31. tabulate age_band, generate (age_dummy)
```

age_band	Freq.	Percent	Cum.
18 and under	**27,627**	**27.99**	**27.99**
19	**24,580**	**24.91**	**52.90**
20	**21,607**	**21.89**	**74.79**
21 and over	**24,877**	**25.21**	**100.00**
Total	**98,691**	**100.00**	

```
32. summarize logApp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
logApp	**98,691**	**3.884657**	**1.488375**	**1.609438**	**8.809863**

```
33. xi: reg logApp Eng i.year CounTrend afterinc male age_dummy2 age_dummy3 age_dummy4 l
  > ogPop, vce(cluster app_country)
  i.year            _Iyear_2008-2015    (naturally coded; _Iyear_2008 omitted)

  Linear regression                               Number of obs     =     98,691
                                                  F(1, 1)           =          .
                                                  Prob > F          =          .
                                                  R-squared         =     0.1908
                                                  Root MSE          =     1.3389

                                (Std. Err. adjusted for 2 clusters in app_country)
```

logApp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Eng	**-28.2329**	**.6579349**	**-42.91**	**0.015**	**-36.59276**	**-19.87305**
_Iyear_2009	**.0558071**	**.0143485**	**3.89**	**0.160**	**-.1265081**	**.2381224**
_Iyear_2010	**.1410544**	**.0147971**	**9.53**	**0.067**	**-.0469608**	**.3290697**
_Iyear_2011	**.1568105**	**.0240297**	**6.53**	**0.097**	**-.1485154**	**.4621364**
_Iyear_2012	**.2954106**	**.0132717**	**22.26**	**0.029**	**.1267782**	**.4640431**
_Iyear_2013	**.3043239**	**.0099662**	**30.54**	**0.021**	**.1776915**	**.4309563**
_Iyear_2014	**.225194**	**.0027244**	**82.66**	**0.008**	**.190577**	**.2598109**
_Iyear_2015	**.2192976**	**.0083178**	**26.36**	**0.024**	**.1136104**	**.3249847**
CounTrend	**.0139867**	**.0012909**	**10.84**	**0.059**	**-.0024151**	**.0303886**
afterinc	**-.2631216**	**.0036969**	**-71.17**	**0.009**	**-.3100956**	**-.2161476**

male	-.0731786	.0138602	-5.28	0.119	-.2492897	.1029324
age_dummy2	-.3257207	1.002878	-0.32	0.800	-13.0685	12.41706
age_dummy3	-1.179068	1.140803	-1.03	0.490	-15.67434	13.3162
age_dummy4	-2.863981	4.553579	-0.63	0.643	-60.72269	54.99473
logPop	.5435658	1.405186	0.39	0.765	-17.31101	18.39814
_cons	1.370985	5.988809	0.23	0.857	-74.72405	77.46602

```
34.
35. *4. STEM and non-STEM subjects
36.
37. gen NOSTEM = STEM==0

38. gen AIncSTEM = afterinc*STEM

39. gen AIncNOSTEM = afterinc*NOSTEM

40. xi: reg logApp Eng i.year i.STEM AIncSTEM AIncNOSTEM male age_dummy2 age_dummy3 age_
  > dummy4 logPop, vce(cluster app_country)
  i.year            _Iyear_2008-2015    (naturally coded; _Iyear_2008 omitted)
  i.STEM            _ISTEM_0-1          (naturally coded; _ISTEM_0 omitted)

  Linear regression                               Number of obs     =     98,691
                                                  F(0, 1)           =          .
                                                  Prob > F          =          .
                                                  R-squared         =     0.1923
                                                  Root MSE          =     1.3378

                                (Std. Err. adjusted for 2 clusters in app_country)
```

logApp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Eng	-.1343815	3.256313	-0.04	0.974	-41.50976	41.241
_Iyear_2009	.067653	.0190684	3.55	0.175	-.1746339	.3099399
_Iyear_2010	.1646766	.0243041	6.78	0.093	-.1441366	.4734899
_Iyear_2011	.1921672	.0383476	5.01	0.125	-.2950849	.6794193
_Iyear_2012	.2953888	.0128591	22.97	0.028	.1319985	.4587791
_Iyear_2013	.3156502	.0058311	54.13	0.012	.2415592	.3897411
_Iyear_2014	.2481778	.006204	40.00	0.016	.1693485	.3270071
_Iyear_2015	.25373	.0056587	44.84	0.014	.1818291	.3256309
_ISTEM_1	-.0025024	.0068401	-0.37	0.777	-.0894135	.0844087
AIncSTEM	-.0993385	.0022343	-44.46	0.014	-.1277281	-.0709488
AIncNOSTEM	-.2788126	.0047135	-59.15	0.011	-.3387033	-.2189219
male	-.0743003	.0134708	-5.52	0.114	-.245463	.0968624
age_dummy2	-.3235274	1.003766	-0.32	0.802	-13.07758	12.43053
age_dummy3	-1.176565	1.141618	-1.03	0.490	-15.6822	13.32907
age_dummy4	-2.876263	4.559694	-0.63	0.642	-60.81267	55.06014
logPop	.5471181	1.406968	0.39	0.764	-17.3301	18.42434
_cons	1.33965	5.986079	0.22	0.860	-74.7207	77.4

```
41.
42. *5.  Expected salary
43.
44. gen AincQ1 = afterinc*Quartile== 1

45. gen AincQ2 = afterinc*Quartile== 2

46. gen AincQ3 = afterinc*Quartile== 3
```

```
47. gen AincQ4 = afterinc*Quartile== 4

48.
49. xi: reg logApp Eng i.year i.Quartile AincQ1 AincQ2 AincQ3 AincQ4 male age_dummy2 age
  > _dummy3 age_dummy4 logPop, vce(cluster app_country)
  i.year            _Iyear_2008-2015    (naturally coded; _Iyear_2008 omitted)
  i.Quartile        _IQuartile_1-4      (naturally coded; _IQuartile_1 omitted)

  Linear regression                               Number of obs     =     67,494
                                                  F(0, 1)           =          .
                                                  Prob > F          =          .
                                                  R-squared         =     0.2937
                                                  Root MSE          =     1.2206

                                (Std. Err. adjusted for 2 clusters in app_country)
------------------------------------------------------------------------------
             |               Robust
      logApp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Eng |  -1.629631   4.335951    -0.38   0.771    -56.72311    53.46385
 _Iyear_2009 |    .082825   .0224205     3.69   0.168    -.2020543    .3677042
 _Iyear_2010 |   .1944411   .0433718     4.48   0.140    -.3566495    .7455318
 _Iyear_2011 |    .212382   .0671333     3.16   0.195    -.6406279    1.065392
 _Iyear_2012 |   .3268919   .0150047    21.79   0.029     .1362389    .5175448
 _Iyear_2013 |   .3697426   .0106682    34.66   0.018     .2341901     .505295
 _Iyear_2014 |   .2970734   .0232492    12.78   0.050     .0016641    .5924827
 _Iyear_2015 |   .3046339   .0207057    14.71   0.043     .0415436    .5677242
_IQuartile_2 |   .0324227   .0175522     1.85   0.316    -.1905987    .2554441
_IQuartile_3 |   .1784818   .0257599     6.93   0.091    -.1488289    .5057925
_IQuartile_4 |   .0767362   .1393897     0.55   0.680    -1.694378     1.84785
       AincQ1 |  -.2475717   .0487972    -5.07   0.124     -.867599    .3724555
       AincQ2 |  -.2461751   .0304491    -8.08   0.078    -.6330674    .1407173
       AincQ3 |  -.2225253   .0212637   -10.47   0.061    -.4927068    .0476562
       AincQ4 |  -.1143015   .0934746    -1.22   0.436    -1.302009    1.073406
         male |  -.0323609   .0365596    -0.89   0.539    -.4968949     .432173
   age_dummy2 |   .1549867   1.394613     0.11   0.930    -17.56525    17.87522
   age_dummy3 |  -.8896616   1.550793    -0.57   0.668    -20.59435    18.81503
   age_dummy4 |  -5.793726   6.026325    -0.96   0.513    -82.36544    70.77799
       logPop |   1.352533   1.861215     0.73   0.600    -22.29644    25.00151
        _cons |  -1.967039   7.865503    -0.25   0.844    -101.9077    97.97365
------------------------------------------------------------------------------

50.
```