ECO 102: Topics in Economics
Ecole Polytechnique, Spring 2022
Professors: Geoffroy Barrows, Benoit Schmutz
Teaching Assistants: Arnault Chatelain, Maddalena Conte

TD 4 (part 2): The marriage market

In this TD we will evaluate the role of cities in the marriage market and assortative matching. As usual, download the folder TD_2_4.zip, write your answers directly in the Tex script and type your Stata commands in a do file.

Exercise

1. We consider the heterosexual marriage market. People's qualities are summarized by a single index. Society is made of three men $\{1,3,5\}$ and three women $\{2,4,6\}$. Couples produce a joint output Z. We consider three different input-output matrices (with males in columns and females in rows).

Case 1

	1	3	5
2	2	6	10
4	4	12	20
6	6	18	30

Case 2

	1	3	5
2	9	15	21
4	15	21	27
6	21	27	33

Case 3

	1	3	5
2	25	39	53
4	41	51	61
6	57	63	69

(a) Assume that output is shared following a competitive bidding process. For each case, guess who will end up marrying whom and explain why.
Correction: Since by the definition, competitive bidding ensures market efficiency which means the sum of the output of all couples in all possibilities is the highest. Therefore we can analyse for each cases by this theory. We denote $(M, W) \in \{1,3,5\} \times \{2,4,6\}$

Case 1: We have the formula of the output $Z(M,W) = MW$ and we know we have the pairs $(1,2)$, $(3,4)$, and $(5,6)$ and overall output $Z = 44$

Case 2: We have the formula of the output $Z(M,W) = 3(M+W)$ since for all posibilities we have the same output and we know we have the possible pairs $\{(1,2),(3,4),(5,6)\}$, $\{(1,2),(3,6),(5,4)\}$ $\{(1,4),(3,2),(5,6)\}$ $\{(1,4),(3,6),(5,2)\}$ $\{(1,6),(3,4),(5,2)\}$ $\{(1,6),(3,2),(5,4)\}$and overall output $Z = 63$

Case 3: We know we have the pairs $(1,6)$, $(3,4)$, and $(5,2)$ and overall output $Z = 161$

(b) Assume now that output is shared according to some predetermined sharing rule. For each case, guess who will end up marrying whom and explain why.
Correction: Since the output is shared and the selection is under some predetermined sharing rule, therefore the people with high index is more likely to marriage with the high index people, so we have in Case 1: $(1,2)$, $(3,4)$, and $(5,6)$ and for Case 2: $\{(1,2),(3,4),(5,6)\}$ and for Case 3: $(1,2)$, $(3,4)$, and $(5,6)$

2. Stata exercise: in this exercise we will explore the question of whether larger cities provide better "marriage markets", due to their higher density.

   (a) Load the data and understand the data structure.

   (b) Summarize education of respondent and partner by city size and export results to Overleaf. Do you notice any pattern?

	mean	sd	min	max	count
0					
education_r	9.74	3.62	5.00	16.50	18,942.00
education_p	9.42	3.71	5.00	16.50	18,942.00
1					
education_r	9.85	3.73	5.00	16.50	12,615.00
education_p	9.39	3.75	5.00	16.50	12,615.00
4					
education_r	9.76	3.86	5.00	16.50	11,046.00
education_p	9.36	3.80	5.00	16.50	11,046.00
6					
education_r	10.55	3.99	5.00	16.50	23,243.00
education_p	10.05	4.00	5.00	16.50	23,243.00
8					
education_r	11.18	4.26	5.00	16.50	12,939.00
education_p	10.77	4.21	5.00	16.50	12,939.00
Total					
education_r	10.23	3.93	5.00	16.50	78,785.00
education_p	9.81	3.93	5.00	16.50	78,785.00
N	78785				

Table 1: The average education years with the different city sizes

Here we have 0 - rural, 1 - small cities, 4 - medium cities, 6 - large cities, 8 - Paris Region.

Therefore we can clearly find that as the growth in city size, the average of the education increase both in education of respondent and partner.

   (c) Do cities support assortative matching in terms of education? Regress partner education on the interaction of respondent education and city size, on age of the respondent, and on whether the respondent is active. Add a year dummy and set the weight using the survey weight. Run the same regression on subperiods of the data (before 2003 and after 2003). Store estimates of the 3 regressions. What pattern do you notice on the interaction term between city size and respondent education? How can you interpret this?
We find that as the growth in city size, the correlation of the education of respondent increase, which implies that cities are more likely support assortative matching in terms of education. However, we can not see a huge difference between the years before or after the year 2003.

```
.         reg education_p city_size##c.education_r age_r active_r i.year [pw=weight]
(sum of wgt is 4.000000013579514)

Linear regression                               Number of obs     =     78,785
                                                F(14, 78770)      =    3777.01
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4408
                                                Root MSE          =     2.9437

------------------------------------------------------------------------------------
                   |               Robust
       education_p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
         city_size |
                 1 |  -.0993127   .1043755    -0.95   0.341     -.303888    .1052626
                 4 |  -.0392729   .1131497    -0.35   0.729    -.2610457    .1824999
                 6 |  -.3634498   .0932717    -3.90   0.000    -.5462617   -.1806379
                 8 |   -.646398   .1091732    -5.92   0.000    -.8603769   -.4324192
                   |
       education_r |   .4724792   .0069714    67.77   0.000     .4588153    .4861431
                   |
city_size#c.education_r |
                 1 |   .0024356   .0105179     0.23   0.817    -.0181795    .0230507
                 4 |   .0111758   .0112773     0.99   0.322    -.0109277    .0332793
                 6 |    .062339   .0089558     6.96   0.000     .0447856    .0798924
                 8 |    .115351   .0097124    11.88   0.000     .0963148    .1343873
                   |
             age_r |   -.052105   .0012913   -40.35   0.000    -.0546359   -.0495741
          active_r |   .4170943   .0438892     9.50   0.000     .3310717    .5031168
                   |
              year |
              2002 |   .7084928   .0314432    22.53   0.000     .6468643    .7701213
              2006 |   .6558956    .035678    18.38   0.000     .5859668    .7258243
              2013 |   .9472047   .0378813    25.00   0.000     .8729577    1.021452
                   |
             _cons |   6.591774   .1158028    56.92   0.000     6.364801    6.818746
------------------------------------------------------------------------------------
```

Figure 1: The regression result of all the years result

(d) We now look at the code to prepare a coefficient plot. The code uses the Stata command *coefplot*. It uses the 3 sets of estimates stored from the previous regression. It then connects the coefficients to show the trend by city size, using the option *recast(connected)*. Look at the code and understand what each options does.

(e) Do cities support assortative matching in terms of employment (i.e. being active on the labor market)? Regress partner active status on the interaction of respondent active status and years, on age and education of the respondent. Set the weight using the survey weight. Run the same regression on different groups of city size (city size less than 5 and city size above 5). Store estimates of the 3 regressions. What pattern do you notice on the interaction term between year and respondent active status? How can you interpret this?

(f) Prepare the coefficient plot as earlier, plotting the coefficients of the interaction of respondent active status and years. Use the 3 sets of estimates stored from the previous regression and drop coefficients not of interest using the option *drop*. Label the coefficients depending on the year. Label estimates from the 3 regressions depending on the city size group. Export the coefficient plot and load it on Overleaf.

```
.         reg education_p city_size##c.education_r age_r active_r i.year [pw=weight] if year<2003
(sum of wgt is 1.999999973979129)

Linear regression                               Number of obs     =     38,686
                                                F(12, 38673)      =    3026.91
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4412
                                                Root MSE          =      2.932

------------------------------------------------------------------------------
                 |               Robust
     education_p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
       city_size |
              1  |  -.0928454   .1176841    -0.79   0.430    -.3235092    .1378184
              4  |    .020764   .1344852     0.15   0.877    -.2428303    .2843583
              6  |  -.3505599    .108371    -3.23   0.001    -.5629697     -.13815
              8  |  -.5574787    .136216    -4.09   0.000    -.8244656   -.2904919
                 |
     education_r |   .4620901   .0086482    53.43   0.000     .4451395    .4790408
                 |
city_size#c.education_r |
              1  |   .0098778   .0126995     0.78   0.437    -.0150136    .0347692
              4  |   .0143461    .014315     1.00   0.316    -.0137116    .0424039
              6  |   .0694198   .0110366     6.29   0.000     .0477878    .0910518
              8  |   .1239333   .0126981     9.76   0.000     .0990448    .1488219
                 |
           age_r |  -.0626662   .0016097   -38.93   0.000    -.0658213   -.0595111
        active_r |   .1193215   .0545525     2.19   0.029     .0123971    .2262458
                 |
            year |
           2002  |   .7286508   .0316797    23.00   0.000     .6665578    .7907438
                 |
           _cons |   7.339842   .1403766    52.29   0.000       7.0647    7.614984
------------------------------------------------------------------------------

.         estimates store r2003
```

Figure 2: The regression result of the years before 2003

```
.         reg education_p city_size##c.education_r age_r active_r i.year [pw=weight] if year>2003
(sum of wgt is 2.000000039600384)

Linear regression                               Number of obs     =     40,099
                                                F(12, 40086)      =    1626.62
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4269
                                                Root MSE          =      2.951

------------------------------------------------------------------------------
                 |               Robust
     education_p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
       city_size |
              1  |   -.166993   .1820218    -0.92   0.359    -.5237599    .1897739
              4  |  -.1747755   .1913699    -0.91   0.361    -.5498648    .2003139
              6  |  -.4517156   .1606482    -2.81   0.005    -.7665898   -.1368413
              8  |  -.8893738   .1800002    -4.94   0.000    -1.242178   -.5365692
                 |
     education_r |   .4760576   .0110976    42.90   0.000      .454306    .4978092
                 |
city_size#c.education_r |
              1  |   .0018898   .0169004     0.11   0.911    -.0312354    .0350151
              4  |   .0155454   .0176624     0.88   0.379    -.0190733     .050164
              6  |   .0631087   .0143269     4.40   0.000     .0350277    .0911897
              8  |   .1203331   .0150881     7.98   0.000     .0907602     .149906
                 |
           age_r |  -.0425236   .0019747   -21.53   0.000     -.046394   -.0386533
        active_r |   .6869731   .0675384    10.17   0.000     .5545963      .81935
                 |
            year |
           2013  |   .2809223    .040376     6.96   0.000     .2017845    .3600601
                 |
           _cons |   6.609607    .190905    34.62   0.000     6.235429    6.983785
------------------------------------------------------------------------------

.         estimates store r2007
```

Figure 3: The regression result of the years after 2003

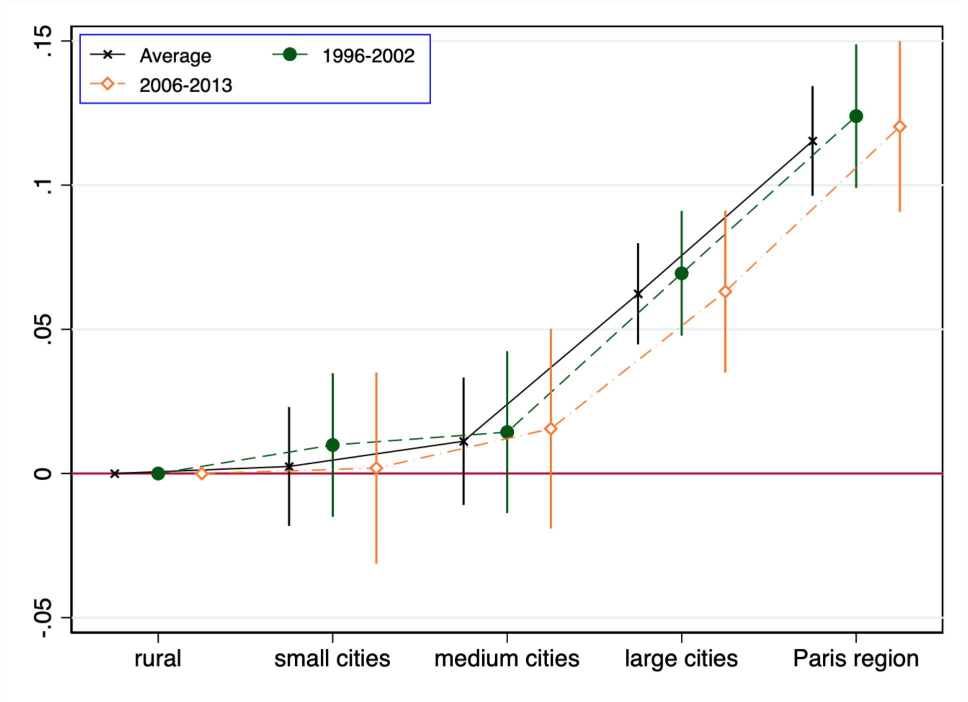

Figure 4: mating graph by the education