

## Research Project description Micro<sup>1</sup>

### “Estimating partner preferences of online daters in high-paid and low-paid occupations”

Marriage partners generally match on intelligence, education, social background, ethnic origin, religion, height, weight, attractiveness, and no doubt other traits too.<sup>2</sup> Why is this? There are a number of explanations. One is a preference story: if individuals prefer marriage partners with traits similar to their own (or prefer traits in partners in similar ways), marriage partners will match on similarities. Another story is one of search frictions: if individuals are more likely to meet their partner in self-selected environments (such as clubs, schools, jobs, neighborhoods), they are also more likely to meet a partner with traits similar to their own, regardless of partner preferences. It has proven difficult, however, to empirically isolate the impact of partner preferences using information on realized partner matches alone.

One possible solution for identifying partner preferences is to run a field experiment in the context of an online dating website and send random invitations from fictitious profiles to online daters. With profiles traits manipulated along two dimensions (attractiveness and education), the responses of online daters will then measure the true preferences for partner attractiveness and education. This is what Egebark, Ekström, Plug, and Van Praag (2021) do in their paper *Brains or Beauty? Causal Evidence on the Returns to Education and Attractiveness in the Online Dating Market*. It is highly recommended that you read this paper.

This project builds on the work of Egebark et al (2021). In particular, students are asked to test whether online daters who work in high-paid occupations respond differently to profile invitations than online daters who work in low-paid occupations. This project will provide insights on a number of theoretical marriage models suggesting that partner preferences might vary with the earnings potential of husbands and wives. In traditional models of household specialization, for example, women are said to prefer men in high-paid occupations over men in low-paid occupations because they can offer more resources for raising children (Becker 1981).

The two data files for this project are (i) a restricted version of the data file used in Egebark et al (2021), with information on the online daters' age, attractiveness, education, gender, occupation (four-digit occupation classification ISCO 2008), type of profile invitation, and reply; (ii) a data file drawn from the Wage Indicator Survey, with information on hourly earnings per occupation (four-digit ISCO 2008).<sup>3</sup>

With these data files, students have to (i) distinguish high-paid (above median) from low-paid (below median) occupations for men and women separately; (ii) replicate the main findings reported in Egebark et al. (2021) Table 5;<sup>4</sup> (iii) test whether main findings hold for a selected sample of employed online daters by estimating the partner preferences for online daters with (non-missing) occupational information; and (iv) test whether the partner preferences are different for male and female online daters working in high-paid and low-paid occupations. Finally, they have to describe these findings in a coherent paper.

The papers and data file are available for download on Blackboard.

---

<sup>1</sup> Developed by Erik Plug.

<sup>2</sup> Economists call these marriage patterns positive assortative matching. Sociologists call these patterns homogamy.

<sup>3</sup> For details on the Wage Indicator Survey, we refer to Tjebens, and Osse. (2015). *Wage Indicator continuous web-survey on work and wages*. Amsterdam: University of Amsterdam/AIAS and Stichting Loonwijzer.

<sup>4</sup> Your results will be slightly different because not all control variables are included in your dataset.

To help you with the analysis, here are some tips:

1. First you will have to merge the two files. The codes in your second file are not unique, you will therefore not use the merge 1:1 but merge m:1. The rest of the syntax is similar (see clab 4.2 Q 5).
2. Which data do you not need anymore and can be dropped?
3. Inspect the data. Which variables are in your data set. Inspect the means for all variables. Specify them by gender and inspect the differences.
4.
  - a. You will have to create (generate) two new variables: a dummy for whether or not the wage is missing, and a dummy for whether or not someone earns above the median. The second part of the syntax of the first dummy is `mi(variable)`.
  - b. In order to find out what the median is type: `tabstat wagehour, by(female) s(median)`. With this information you will be able to create this dummy. Do this in two steps: 1. Generate the variable and write above which number the wage is consider high and end the statement with “ `if variable==0 & wage_missing==0`”. Step 2: use the replace command to fill in the high wages of the other gender. Should you set `wage_missing` to zero or to one? Think carefully!
5. You might want to create tables in which all models are summarized together. To do that, type first *ssc install estout*. With the `eststo` command you can indicate which models you want to put in one table (Cf. Clab 4.1 or clab4.2).
6. Check first whether the three groups are equal on all variables. Use a regression. What should be the dependent variable? Above the table you will find the probability (p-value) that the model is significant: `Prob>F`. If this p-value of each of the models is above 0.05, the model is not significant. This means that the dependent variable is not predicted by the independent variables.
7. If you want to do a regression for only women (or men) or high paid workers use ‘ `if 'variable'==1 (or 0)`’ or ‘ `if 'variable1'==1 (or 0) & variable2==1 (or 0)`’ at the end of the regression. This way you tell STATA only to include the variable that fulfill this criterion. Mind though that you still have to decide which variable you will set to one or zero. Replicate the regressions from table 5 on p.21 of the article.
8. To check whether there is an interaction effect of education and attractiveness create an interaction variable. This interaction term will have the following form: `c.variable1#c.variable2`. You have to fill in a variable name for 'variable 1' and 'variable2'. See also clab4.1. How many interaction variable should you make? Check the variable list again. You will need this information for question ii.
9. Check the partner preferences for online daters with non-missing occupational information
10. Check whether male and female online daters working in high-paid and low-paid occupations have different partner preferences.

**Key reference**

[Egebark, Ekström, Plug, and Van Praag. \(2021\). Brains or Beauty? Causal Evidence on the Returns to Education and Attractiveness in the Online Dating Market. \*Journal of Public Economics\* 196.](#)

**Related references**

Becker. (1981). *A Treatise on the Family*. Cambridge, MA: Harvard University Press

[Fisman, Iyengar, Kamenica, and Simonson. \(2006\). Gender differences in mate selection: Evidence from a speed dating experiment. \*Quarterly Journal of Economics\* 121, 673-697.](#)

[Hitsch, Hortacsu, and Ariely. \(2010\). Matching and sorting in online dating. \*American Economic Review\* 100, 130–163.](#)