Read Me File - Replication Package

The Role of the Ask Gap in Gender Pay Inequality

Nina Roussille

December, 2023

Proprietary Data

The Hired.com data is proprietary and therefore not made available to the public. Hence the Data folder of this replication package in empty. However, this Read me file details how each dataset was used to produce the tables and figures in the paper.

Overview of the code

In the Code folder, you will find all .do and .R files that create the figures and tables shown in the paper "THE ROLE OF THE ASK GAP IN GENDER PAY INEQUALITY". After adapting the global file paths, running Master.do will produce all the figures and tables for the main paper and the appendix (but Figure C1 and Table C1, which are coded in R).

1. **Master.do**: Runs the do-files listed below.
2. **Globals.do**: Defines globals for loading data, controls, and the creation of tables.
3. **Main_Tables.do**: Creates Paper Tables.
4. **Main_Figures.do**: Creates Paper Figures.
5. **Appendix_Output.do**: Creates Online Appendix Figures and Tables (but Figure C1 and Table C1, that are coded in R).
6. **Heterogeneity.R**: Creates Figure C1 and Table C1.

1 Description of each Do-Files

1.1 Master.do

This do-file installs the required packages, specifies the data paths, and runs the do-files listed thereafter. The global paths need to be adapted to one's local directory.

Running the master do-file suffices to produce all the data-generated Figures and Tables in the main paper, as well as all but Figure C1 and Table C1 in the Online Appendix. The Heterogeneity.R file needs to be run separately to get Figure C1 and Table C1.

1.2 Construction.do

This do-file uses as input all the raw data files shared by Hired.com, cleans, and merges them into a single, processed .dta file that is used in the rest of the analysis. Most variables used for the analysis are also labeled at the end of the construction file so that the reader can refer to that as a dictionary for the variables used in the Main_Tables.do and Main_Figures.do and Appendix_Output.do.

1.2.1 Data Input

- **compensation.csv**: this file contains details on whether a given bid also had an equity offer.
- **matched_education.csv**: this file, with observations at the candidate level, contains all the education characteristics of candidates, as described in Online Appendix Table A1.
- **most_recent_company.csv**: this file, with observations at the candidate level, contains the most recent company a candidate worked at when they received a given bid. This file is not exactly a raw file: it is the result of a cleaning of company names as described in Online Appendix Section I.4. The code for the cleaning was deemed proprietary by Hired.com since it contained raw company names, which is why I did not make it available in the replication package. Each step of the cleaning is however described in Online Appendix Section I.4.
- **company_id_benefits.csv**: this file, with observations at the company level, contains the non-wage benefits fields that firms fill out on Hired.com.
- **platform_companies_info.csv**: this file, with observations at the company level, contains some basic firm characteristics (e.g. firm size) that firms list on their Hired.com profile.

- **jobtitles.csv**: this file, with observations at the candidate level, contains the successive job titles candidates held at their previous job(s).
- **top five skills.csv**: this file, with observations at the candidate level, contains the (top five) skills that candidates list on their Hired.com profile.
- **candidate_wishlist_data.csv**: this file, with observations at the candidate level, contains the next job wishes (e.g. type of position) that candidates list on their Hired.com profile when they receive a given bid.
- **candidate_offers.csv**: this file, with observations at the bid level, contains details on every single bid firms sent (so one line is at the job × candidate level). For each candidate (× job) observation, the file also contains most of (the rest are in separate files described in the next bullet points) the resume characteristics of the individual at the time the bid was made (as described in Online Appendix Table A1), the ask, bid and (if it applies), final salaries, as well as a company identifier. For candidates that did not receive any bids, this file still contains one observation per candidate (with a missing job id), providing information on their resume.

1.2.2 Data Output

- **SGenderMetaD.dta**: the construction do-file describes in detail how I get to this dataset. It is basically a version of candidate_offers augmented by all the other datasets, with some additional cleaning and variables created for the purpose of the paper.

1.3 Globals.do

This do-file defines globals for loading data, for controls, and for the creation of tables.

1.4 Main_Tables.do

This do-file creates all the data-generated Tables as shown in the paper.

1.4.1 Data Input

- **SGenderMetaD.dta**
- **unique_ranking.csv**: This file provides each Hired.com company's non-wage amenity value and expected match productivity with a candidate, based on estimates in Roussille and Scud-

eri (2023).

1.5 Main_Figures.do

This do-file creates all the data-generated Figures as shown in the paper.

1.5.1 Data Input

- **SGenderMetaD.dta**

1.6 Appendix_Output.do

This do-file creates figures and tables appearing in the paper's Online Appendix.

1.6.1 Data Input

- **SGenderMetaD.dta**
- **unique_ranking.csv**: This file contains each Hired.com company's non-wage amenity value, based on estimates in Roussille and Scuderi (2023).
- **prod_ask.dta**: This file (at the candidate × company id level) contains Hired.com company's expected match productivity with a given candidate, based on estimates in Roussille and Scuderi (2023).

2 R File

2.1 Heterogeneity.R

This file creates Figure C1 and Table C1, from the paper's Online Appendix Section C.

2.1.1 Data Input

- **SGenderMetaD.dta**

References

Roussille, Nina, and Benjamin Scuderi. 2023. "Bidding for Talent: A Test of Conduct on a High-Wage Labor Market." *Working paper.*