Gender Gap in Perceived vs Actual Ability: Evidence from a Large-Scale Labour Market Experiment

Jan Feld (Victoria University of Wellington), Edwin Ip (University of Exeter), Andreas Leibbrandt (Monash University), Joseph Vecci (University of Gothenburg)

We employ a novel approach to conduct a large-scale labour market experiment that can measure the actual ability of job applicants as well as the employers' perception of their ability. In this pre-registered field experiment, we focus on the gender differences in actual and perceived abilities of real job applicants in the context of the technology sector.

The experiment consists of two stages. In the first stage, we measure the job applicant's actual ability. We advertised for a real python programmer job on all major job sites in the United States. Of the 1200 applicants, we invited a stratified random sample of eligible applicants to complete standardised programming tasks in the second stage of their application in order to measure their on-the-job skills. In total, 72 female applicants and 254 male applicants completed the tasks. We comprehensively evaluate the applicants' coding ability using their codes from the programming tasks. To ensure the objectivity and validity of our skill measurement, the programming tasks and its evaluation were carefully designed and selected using an expert survey of 8 programmers that we recruited separately. Applicants' codes are holistically evaluated according to five criteria (test case score, efficiency, complexity, style and frequency of errors) using a bespoke script written by 2 professional programmers that we hired independently. The overall score (out of 100

points) is calculated using a weighted average of the five criteria determined by the expert survey.

In the second stage, we elicit employers' beliefs about the ability of the applicants. We recruit 240 programmers and HR professionals who are involved in hiring programmers in the United States as participants ("employers"). We provide them with information about the job advertisement, the programming tasks and the evaluation method. We then ask the employers to each guess 10 applicants' programming scores based on basic information from the applicants' CV (first name, education, years of experience, etc.). Employers are incentivised based on how close their guesses are to the candidate's actual score using a quadratic scoring rule. The closer their guess is to the actual score, the higher their chance of getting a large monetary bonus. Finally, we use incentivised procedures to ask employers their beliefs about the distribution of skills by gender both within the applicant sample and in the general population, which allows us to explore potential mechanisms.

We find that there are no differences between male and female applicants in their actual coding skills. Despite this, employers believe that female applicants are significantly less skilled than male applicants. All else equal, employers believe that female applicants' score in the programming task is 8.8 points (or 13.3%) worse than their male counterparts (p=0.00). This difference corresponds to 0.25 standard deviation. We found that HR professionals are much more likely to have this false belief than professional programmers. The results do not differ by the gender of the employers. Our results cannot be explained by representativeness heuristics or

attention discrimination, but they could be partially explained by employers neglecting the selection effect: that gender difference in programming ability is likely to be minimal among those who select into a programming career compared to those in the general population.

Our study provides the first field evidence on inaccurate beliefs about gender differences in productivity, which forms the basis for inaccurate statistical discrimination, an idea that has received increasing attention recently in the economics literature.