Pre-Analysis Plan for "Understanding Employee Sorting between Startups"

Kevin Bryan Mitchell Hoffman Amir Sariri

August 24, 2019

1 Introduction

This project examines why workers sort between different startups, focusing on the role of expert information on different aspects of firm quality. Unlike for established firms, where perspective workers have various sources of information (e.g., best places to work rankings or online reviews), workers may have relatively little information about different startups. In addition, workers may have a hard time evaluating the technology or business model of startups, either because they lack very specialized technical training or because of firm secrecy. In this project, we examine how expert opinions about the business and science quality of startups affect worker demand for working at particular startups. Beyond the overall treatment effects of expert information, we are interested in the mechanisms for the treatment effects.

The main idea of the experimental design is to take workers who are potentially interested in working for startups, and then shock them with information about the science quality and/or business quality of firms. The experiment is being carried out by a leading science-based entrepreneurship program (SEP) that focuses on scalable pre-seed startups. The experiment will use a 2x2 experimental design, where the arms are expert information about science quality and expert opinion about business quality. Business and science experts rate science-based startups in terms of business quality or science quality. Simple information from their ratings (e.g., above-average or not) is then communicated to job candidates, depending on the treatment group. That is workers are assigned to one of four groups: Control, Business Quality Only, Science Quality Only, or Both.

1.1 Experimental Plan

As of now (August 2019), we have conducted preliminary analysis of the MBA RCT, for which we have unincentivized interest data but no incentivized interest data. However, we have not analyzed data from the RCT with business program alum, for which there is incentivized ranking data.

2 Pre-Analysis Plan

The key questions in this study are:

1. Do expert opinions on science or business quality affect worker interest in working for different startups? This is the most important question.

2. Do expert opinions on science or business quality affect workers' perceptions of startups' business and science quality?

3. Do expert opinions affect perceived probabilities of firm success? Are such belief a mechanism for the impact of expert opinions on worker interest in different startups?

2.1 Primary Outcome Variables

Given our interest in understanding how information affects interest in working at firms, our main outcomes are:

- Unincentivized interest in working at a company. This is based on a scale from 1-5, but we will analyze in normalized form.

- Incentivized ranking of companies. Subjects are allowed to rank up to 10 companies. We intend to analyze several different outcomes, including (1) a dummy for whether company is ranked at all; (2) a dummy for whether a company was ranked as a person's top choice; (2) a dummy for whether a company was ranked in a person's top 3 choices; and (4) a ranking variable that imputes unranked companies to a lower rank. For (4), we analyze it in both normalized form (to get a better sense of magnitudes) and un-normalized form.

  As a robustness check, we will also perform the analysis only using companies that are ranked. We will reverse code the ranking variables so that a higher ranking indicates a more preferred company.

2.2 Secondary Outcome Variables

The secondary outcome variables help us understand whether/how the treatments affect workers' beliefs about the firms.

- Perceived quality of the science. This is based on a scale from 1-5, but we will analyze in normalized form.
- Perceived quality of the business. This is based on a scale from 1-5, but we will analyze in normalized form.
- Perceived probability of raising money at a $1 million valuation within one year.
- Perceived probability of achieving either an initial public offering (IPO) or acquisition at a price of more than $50 million within one year.
- Perceived probability of a firm being selected to graduate from the science-based entrepreneurship program (SEP). This outcome will only be asked in some surveys.

2.3 Inclusion/Exclusion

A small number of data points are missing since a given scientist did not submit the survey in time.

2.4 Statistical model specifications and hypotheses

We are interested in whether providing information has a differential positive effect when the information is positive. The regressions we analyze will be of the form:

$$\begin{aligned} y_{nf} &= a_0 + a_1 GotBizInfo_n + a_2 GotBizInfo_n * GoodBizFirm_f + X_{nf} + \varepsilon_{nf} \\ y_{nf} &= b_0 + b_1 GotScienceInfo_n + b_2 GotScienceInfo_n * GoodScienceFirm_f + X_{nf} + \varepsilon_{nf} \end{aligned}$$

Here, the index n denotes workers (or subjects) and f denotes a firm that a worker is evaluating. Thus, an observation is a worker-firm. The outcome, y_{nf}, will be one of the above outcomes. The regressor $GotBizInfo_n$ measures whether subject n is randomly assigned to receive information about the business quality of the firm. Likewise, $GotScienceInfo_n$ measures whether a subject randomly receives information about the science quality of the firm. The variable $GoodBizFirm_f$ indicates whether firm f is rated positively or not in terms of business quality, whereas $GoodScienceFirm_f$ indicates whether a firm is rated positively or not in terms of science quality. The controls X_{nf} will include firm fixed effects to control for the underlying quality of the firm, as well as anything else about how the firm is presented (such as the firm's position on the job board). The controls will also include any strata dummies for RCTs when we do a stratified randomization. We will cluster the standard errors by worker to account for the fact that the treatments are at the worker level.

We will examine the null hypotheses of $a_1 = 0$, $a_2 = 0$, and $a_1 + a_2 = 0$, as well as the null hypotheses that $b_1 = 0$, $b_2 = 0$, and $b_1 + b_2 = 0$. We hypothesize that $a_1 < 0$ and $b_1 < 0$; that $a_2 > 0$ and $b_2 > 0$; and that $a_1 + a_2 > 0$ and $b_1 + b_2 > 0$ for the primary and secondary outcomes we analyze. The hypothesis of $a_1 < 0$ means that receiving negative business information makes workers less interested in working a company relative to not receiving such information, conditional on the firm. For the secondary outcomes, we expect they will be negative, particularly on perceptions of business quality. The hypothesis of $b_1 < 0$ means that receiving negative science information makes workers less interested in working a company relative to not receiving such information, conditional on the firm. For the secondary outcomes, we expect they will be negative, particularly on perceptions of science quality.

The hypotheses of $a_2 > 0$ and $b_2 > 0$ means that subjects will be more interested in working at companies when the information they receive is positive about a company compared to when the information they receive is negative about a company.

The hypotheses of $a_1 + a_2 > 0$ and $b_1 + b_2 > 0$ means that receiving positive information makes workers more interested in working at a company relative to not receiving such

information, conditional on a firm.

The hypotheses can also be stated verbally as follows:

H1 (corresponding to $a_1 < 0$): Compared to not receiving information, receiving negative business information makes someone have lower interest in working at a company (incentivized or non-incentivized); have lower perception of business and science quality; and believe the firm will have lower chance of firm success (measured in terms of perceived probability of obtaining a $1 million valuation, a successful exit (IPO or $50 million acquisition), or graduating from the SEP).

H2 (corresponding to $b_1 < 0$): Compared to not receiving information, receiving negative business information makes someone have lower interest in working at a company; have lower perception of business and science quality; and believe the firm will have lower chance of firm success.

H3 (corresponding to $a_2 > 0$): Compared to receiving negative information, receiving positive business information makes someone have higher interest in working at a company; have higher perception of business and science quality; and believe the firm will have higher chance of firm success.

H4 (corresponding to $b_2 > 0$): Compared to receiving negative information, receiving positive science information makes someone have higher interest in working at a company; have higher perception of business and science quality; and believe the firm will have higher chance of firm success.

H5 (corresponding to $a_1 + a_2 > 0$): Compared to not receiving information, receiving positive business information makes someone have lower interest in working at a company; have lower perception of business and science quality; and believe the firm will have lower chance of firm success.

H6 (corresponding to $b_1 + b_2 > 0$): Compared to not receiving information, receiving positive science information makes someone have lower interest in working at a company; have lower perception of business and science quality; and believe the firm will have lower chance

of firm success.

In addition to analyzing these regressions separately, we will also run regressions that include all terms:

$$\begin{aligned} y_{nf} &= c_0 + c_1 GotBizInfo_n + c_2 GotBizInfo_n * GoodBizFirm_f + c_3 GotScienceInfo_n \\ &\quad + c_4 GotScienceInfo_n * GoodScienceFirm_f + X_{nf} + \varepsilon_{nf} \end{aligned}$$

For our primary outcome variable of 1-5 interest in a company, this can be run on all our data pooled together. However, for the incentivized ranking data, this can only be run on people participating in the job board, where the incentivized rankings were elicited.

We will also consider IV regressions of the below form:

$$\begin{aligned} y_{nf} &= \alpha_0 + \alpha_1 PerceivedScienceQuality_{nf} + \alpha_2 PerceivedBizQuality_{nf} + X_{nf} + \varepsilon_{nf} \\ PerceivedScienceQuality_{nf} &= \beta_0 + \beta_1 GotBizInfo_n + \beta_2 GotBizInfo_n * GoodBizFirm_f + \beta_3 GotScienceInfo_n \\ &\quad + \beta_4 GotScienceInfo_n * GoodScienceFirm_f + X_{nf} + \varepsilon_{nf} \\ PerceivedBizQuality_{nf} &= \gamma_0 + \gamma_1 GotBizInfo_n + \gamma_2 GotBizInfo_n * GoodBizFirm_f + \gamma_3 GotScienceInfo_n \\ &\quad + \gamma_4 GotScienceInfo_n * GoodScienceFirm_f + X_{nf} + \varepsilon_{nf} \end{aligned}$$

We are also interested in the level of worker beliefs, in order to test the idea that workers may overestimate the probability of firm success.

2.5 Heterogeneity

Worker heterogeneity. We will examine heterogeneity based on whether workers have an undergraduate degree in STEM or not. The idea is that workers with an undergraduate degree in STEM might be able to better judge the quality of the startups, particularly the science quality.

Firm heterogeneity. Our main regressions already examine heterogeneity based on whether the firm is above- or below-average in terms of science or business quality. In addition, one might imagine that our treatment effects would be larger in more technologically sophisticated industries. While most of the firms we are analyzing are in technologically sophisticated areas, we may also try to perform heterogeneity analysis splitting by whether the sector is particularly technologically sophisticated.