An Audit Alternative: Measuring Employer Preferences and Beliefs without Deception

Judd B. Kessler, Corinne Low, and Colin D. Sullivan

July 23, 2018
What matters on the job market?

- Fundamental question in labor economics is how employers value different candidate characteristics, such as:
  - Human capital characteristics (education, field of study, experience, e.g., Autor and Houseman [2010], Pallais [2014])
  - Gender and race (e.g., Altonji and Blank [1999])

- We need powerful tools to study these questions—obviously observational studies are insufficient

- Audit studies have been a workhorse in this literature
  - In-person (critiqued by Turner et al. [1991], Heckman and Siegelman [1992], Heckman [1998])
  - Correspondence and resume audits for discrimination (large literature launched by Bertrand and Mullainathan [2004])
  - Branched out into new areas (e.g., unemployment spells, Kroft et al. [2013], Eriksson and Rooth [2014], Nunley et al. [2017], value of for-profit-college degrees, Deming et al. [2016])

- Resume audit studies give you the difference in callback rates between groups
Callback indicates a candidate is above a threshold

Imagine a distribution of employer i’s expected productivity of candidate j with vector of characteristics $X_j$ as in the below:

$$V_{ij} = \beta X_j + \xi_{ij},$$
Callback indicates a candidate is above a threshold

What we observe in an audit study is an indicator for whether a candidate is called back:

\[ D_{ij} = 1[V_{ij} \geq V_i^*(c_i)] \]
Callback indicates a candidate is above a threshold

We can compare callback rates of different groups. For simplicity, imagine a binary characteristic $x_j$:

$$V_{ij} = \beta x_j + \xi_{ij},$$
Callback indicates a candidate is above a threshold

- Resume audit studies measure the impact of $x_j$ on callback rate by estimating $\alpha$ as:

$$\alpha = E[D_{ij}|x_j = 1] - E[D_{ij}|x_j = 0]$$
Callback indicates a candidate is above a threshold

- Why might we be interested in richer information on $V_{ij}$?
- If the shape of the distribution depends on $x_j$, callback rates will not have consistent relationship across the distribution.

![Graph showing the distribution of values with a shaded area representing 6.68%]
Callback indicates a candidate is above a threshold

- Why might we be interested in richer information on $V_{ij}$?
- If the shape of the distribution depends on $x_j$, callback rates will not have consistent relationship across the distribution.
Callback indicates a candidate is above a threshold

- Why might we be interested in richer information on $V_{ij}$?
- If you change thresholds (e.g., selective hiring to economic expansion), the sign of $\alpha$ can flip
Callback indicates a candidate is above a threshold

- Why might we be interested in richer information on $V_{ij}$?
- If you change thresholds (e.g., selective hiring to economic expansion), the sign of $\alpha$ can flip

![Graph showing density and value distribution]
A new approach: Incentivized Resume Rating (IRR)

- Much richer information by being able to directly measure preferences
  - Parallel to buy / no-buy versus tracing demand curve (e.g., BDM)

- How to do it in the hiring domain with incentives?
  - Employers rate **hypothetical resumes** with randomly assigned characteristics
  - They are matched with **real job seekers** according to their reported preferences

- Similar-in-spirit to design applied to dating markets in Low [2017]

- This offers the control of a laboratory experiment with the “stakes” of a field experiment
  - Independently randomize many characteristics
  - Get continuous measures of employer preferences
  - Have each employer rate multiple resumes

- Experimental paradigm is very flexible, and can be used to measure many different traits with different pools of employers and candidates
We study employer preferences for college students

- Traditionally hard to investigate preferences of “elite” employers because they do not accept cold resumes

- How they value human capital investments
  - College students spend three months a year outside of school; we explore the impact of their HC accumulation in those months
  - Investigate impact of quality (e.g., more prestigious internship) and quantity (e.g., an additional experience) of summer employment
  - Can compare these to impact of GPA, which we treat as a numeraire

- How they respond to demographics
  - On-campus recruiters may have different race and gender preferences than firms traditionally targeted in resume audit studies
  - We measure—for the first time—employers’ beliefs about demographic groups’ likelihood of job acceptance
Nathan Stewart

EDUCATION

University of Pennsylvania, College of Arts and Sciences
BA in Economics
Cumulative GPA: 3.82/4.00

Philadelphia, PA
Expected May 2017

WORK EXPERIENCE

Bank of America Merrill Lynch, New York, NY
Investment Banking Summer Analyst - Healthcare Finance

- Advised hospitals and healthcare systems on strategic financing options, new project initiatives, and M&A opportunities
- Prepared client pitchbooks and presentation materials for investor roadshows
- Conducted due diligence and filed reports on 103 deals (41% of group's total deals) as part of an SEC initiative (MCDC)

June - August 2016

P.F. Chang's, Mclean, VA
Server

- Memorized entire menu and completed server training in five days
- Worked diligently under stressful conditions to deliver high quality service to customers
- Communicated and worked with servers, hosts, and bar staff to operate restaurant smoothly and uphold P.F. Chang's core values and principles

June - August 2015

LEADERSHIP EXPERIENCE

MUSE - Undergraduate Marketing Club, Philadelphia, PA
Executive Board Member

2013-2015

- Assisted in organizing speaker conferences, alumni panels, and networking sessions, with past
Incentivized Resume Rating: our design

- We partner with University of Pennsylvania Career Services
  - Collect hundreds of real Penn resumes to cull components
  - Use real Penn seniors interested in being matched as candidate pool

- Career Services offers employers the opportunity to try a new pilot tool designed by Wharton professors
  - Framed and marketed as a way to help employers find candidates
  - Only participation incentive is to be matched with Penn seniors

- Employers rate 40 resumes (median employer takes 28 minutes)
  - Choose majors to view: Humanities/Social Sciences or Science/Math
  - Rate candidates on: “desirability” and “likelihood of acceptance”

- We use ML to match each employer to 10 real seniors based on their preferences (i.e., no deception) and email their resumes

- We repeat the experiment at University of Pittsburgh to show differences based on subject pool
Rating on two dimensions

**MKT and EVOLUTION**

**Prospect Forum**, Philadelphia, PA  
Executive VP  
2013-2015

- Launched early stage entrepreneurial venture with peers to improve the career search process for college students
- Connected students with Philadelphia-based companies that match their interests and worked to expand to other Ivies

**SKILLS**

Public speaking, marketing, writing, fundraising, data analysis, PowerPoint, Excel

---

How interested would you be in hiring **Nathan Stewart**?

<table>
<thead>
<tr>
<th>Not interested</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Very interested</th>
<th>10</th>
</tr>
</thead>
</table>

How likely do you think **Nathan Stewart** would be to accept a job with your organization?

<table>
<thead>
<tr>
<th>Not likely</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Very likely</th>
<th>10</th>
</tr>
</thead>
</table>
## Resume creation and variables

<table>
<thead>
<tr>
<th>Component</th>
<th>Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPA</strong></td>
<td>Drawn from $U(2.90, 4.00)$</td>
</tr>
<tr>
<td>Major</td>
<td>Drawn from a list of Penn majors</td>
</tr>
<tr>
<td>First job</td>
<td>$\Pr(\text{Top Internship}) = \frac{1}{2}$</td>
</tr>
<tr>
<td>Second job</td>
<td>$\Pr(\text{Second Internship}) = \frac{13}{40}$, $\Pr(\text{Work for Money}) = \frac{13}{40}$, $\Pr(\text{Blank}) = \frac{14}{40}$</td>
</tr>
<tr>
<td>Leadership</td>
<td>Two items drawn independently</td>
</tr>
<tr>
<td>Skills</td>
<td>$\Pr(\text{Technical skills}) = 0.25$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>$\Pr(\text{Not White Male}) = 67.2%$, Gender (50% Male, 50% Female), Race drawn from U.S. distribution (65.7% White, 16.8% Hispanic, 12.6% Black, 4.9% Asian)</td>
</tr>
<tr>
<td>Race</td>
<td>drawn from U.S. distribution (65.7% White, 16.8% Hispanic, 12.6% Black, 4.9% Asian)</td>
</tr>
</tbody>
</table>
Regression specification

- Recall expected employer productivity, \( V_{ij} = \beta X_j + \xi_{ij} \)

- Mean value in OLS (averaged over the space we created):

\[
V_{ij} = \beta_0 + \beta_1 \text{GPA} + \beta_2 \text{TopInt} + \beta_3 \text{SecondInt} + \beta_4 \text{WFM} + \\
\beta_5 \text{TechSkills} + \beta_6 \text{NotWhiteMale} + \alpha_i + \gamma_j + \xi_{ij}
\]

where \( \alpha_i \) are rater fixed effects and \( \gamma_j \) includes leadership and major fixed effects

- Will also run quantile specifications to estimate marginal effects at 25th, 50th, 75th, 90th, and 95th percentiles

- Will first present results on the first rating: “How interested would you be in hiring [name]”? 
### OLS results

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Humanities &amp; Social Sciences</th>
<th>Science &amp; Math</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPA</strong></td>
<td>2.195***</td>
<td>2.300***</td>
<td>1.852***</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.153)</td>
<td>(0.243)</td>
</tr>
<tr>
<td><strong>Top Internship</strong></td>
<td>0.902***</td>
<td>1.039***</td>
<td>0.530***</td>
</tr>
<tr>
<td></td>
<td>(0.0806)</td>
<td>(0.0944)</td>
<td>(0.173)</td>
</tr>
<tr>
<td><strong>Second Internship</strong></td>
<td>0.463***</td>
<td>0.514***</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td>(0.0947)</td>
<td>(0.114)</td>
<td>(0.187)</td>
</tr>
<tr>
<td><strong>Work for Money</strong></td>
<td>0.149</td>
<td>0.114</td>
<td>0.319*</td>
</tr>
<tr>
<td></td>
<td>(0.0913)</td>
<td>(0.109)</td>
<td>(0.185)</td>
</tr>
<tr>
<td><strong>Technical Skills</strong></td>
<td>-0.0680</td>
<td>-0.0492</td>
<td>-0.171</td>
</tr>
<tr>
<td></td>
<td>(0.0900)</td>
<td>(0.106)</td>
<td>(0.186)</td>
</tr>
<tr>
<td><strong>Not White Male</strong></td>
<td>-0.117</td>
<td>-0.0110</td>
<td>-0.399**</td>
</tr>
<tr>
<td></td>
<td>(0.0842)</td>
<td>(0.0998)</td>
<td>(0.188)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>2880</td>
<td>2040</td>
<td>840</td>
</tr>
<tr>
<td><strong>F-test p-value for Majors</strong></td>
<td>&lt; 0.001</td>
<td>0.0036</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><strong>F-test p-value for Leadership</strong></td>
<td>0.0649</td>
<td>0.0246</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Valuation of summer work experience

Bars indicate 95% Confidence Intervals.
Interactions between work experience

Bars indicate 95% Confidence Intervals.
Work experience narrative?

Bars indicate 95% Confidence Intervals.

- **Top Internship**
- **Top Internship x Any Second Job**
- **Any Second Job**
- **No Work Narrative**
Human capital matters, what about demographics?

- Have shown that firms value summer work experience
  - Both quality and quantity important—effects differ by quantile
  - Constraints students face in needing to earn money from summer work might be materially important
  - Interactions between different components, can be more closely examined with this design

- Have also shown that firms recruiting in STEM are less interested in female/minority candidates
  - Will now examine impact of demographic characteristics more closely
  - In Bertrand and Mullainathan [2004], not only did resumes with black names receive fewer callbacks, there was also a lower return to quality improvements
Top Internship less valuable for women and minorities

Bars indicate 95% Confidence Intervals.
Effect absent for GPA

Bars indicate 95% Confidence Intervals.
Second measure: likelihood of acceptance

- Recall question: “How likely do you think [name] would be to accept a job with your organization?”
  - This is correlated positively with desirability rating
  - Holding desirability constant, negatively correlated with “objective” quality

![Graph showing the likelihood of acceptance with GPA](chart.png)

\[ \beta = -0.94, \ SE = 0.09 \]
Firms believe women and minorities are less likely to accept

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Desirability &lt; 5</th>
<th>Desirability ≥ 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>0.734***</td>
<td>-0.341**</td>
<td>-0.133</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.140)</td>
<td>(0.144)</td>
</tr>
<tr>
<td>Top Internship</td>
<td>0.666***</td>
<td>0.435***</td>
<td>0.0632</td>
</tr>
<tr>
<td></td>
<td>(0.0763)</td>
<td>(0.0910)</td>
<td>(0.0880)</td>
</tr>
<tr>
<td>Second Internship</td>
<td>0.393***</td>
<td>0.293***</td>
<td>0.194*</td>
</tr>
<tr>
<td></td>
<td>(0.0910)</td>
<td>(0.105)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>Work for Money</td>
<td>0.200**</td>
<td>0.0895</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>(0.0895)</td>
<td>(0.0991)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>Technical Skills</td>
<td>-0.105</td>
<td>0.00508</td>
<td>-0.119</td>
</tr>
<tr>
<td></td>
<td>(0.0862)</td>
<td>(0.0982)</td>
<td>(0.0962)</td>
</tr>
<tr>
<td>Not White Male</td>
<td>-0.197**</td>
<td>-0.0664</td>
<td>-0.208**</td>
</tr>
<tr>
<td></td>
<td>(0.0805)</td>
<td>(0.0913)</td>
<td>(0.0919)</td>
</tr>
<tr>
<td>Observations</td>
<td>2880</td>
<td>1367</td>
<td>1513</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p < 0.10, ** p < 0.05, *** p < 0.01
Why does this matter?

- Imagine the firm incurs costs to interview or recruit candidates (e.g., time/effort, limited slots)

- Could produce (or exacerbate) lower callback rates for under-represented groups

- Callback differences may reflect more than expected productivity
  - Essentially an omitted variable bias problem
  - But not solved with randomization, since appeal of trait and impact on likelihood of acceptance assigned simultaneously
  - Anything the firm finds appealing might also change their chance of “getting” candidate
Incentivized Resume Rating: future research opportunities

- IRR can be used to answer a wide array of human capital questions
- Can identify different dimensions of preferences
- Setup costs are substantial, but marginal costs of running are lower (we will gladly share our technology)
- Can be used outside of college setting
- Deployment with multiple groups possible for comparison
Firm Size & GPA

Bars indicate 95% Confidence Intervals.
Firm Size & Top Internship

Bars indicate 95% Confidence Intervals.
Results at Pitt directionally similar

<table>
<thead>
<tr>
<th></th>
<th>Penn</th>
<th>Pitt</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>2.195***</td>
<td>0.263**</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>Top Internship</td>
<td>0.902***</td>
<td>0.222***</td>
</tr>
<tr>
<td></td>
<td>(0.0806)</td>
<td>(0.0741)</td>
</tr>
<tr>
<td>Second Internship</td>
<td>0.463***</td>
<td>0.212**</td>
</tr>
<tr>
<td></td>
<td>(0.0947)</td>
<td>(0.0844)</td>
</tr>
<tr>
<td>Work for Money</td>
<td>0.149</td>
<td>0.154*</td>
</tr>
<tr>
<td></td>
<td>(0.0913)</td>
<td>(0.0807)</td>
</tr>
<tr>
<td>Technical Skills</td>
<td>-0.0680</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>(0.0900)</td>
<td>(0.0768)</td>
</tr>
<tr>
<td>Not White Male</td>
<td>-0.117</td>
<td>0.00297</td>
</tr>
<tr>
<td></td>
<td>(0.0842)</td>
<td>(0.0710)</td>
</tr>
<tr>
<td>Observations</td>
<td>2880</td>
<td>3440</td>
</tr>
<tr>
<td>F-test p-value for Majors</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>F-test p-value for Leadership</td>
<td>0.0649</td>
<td>0.937</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
In this paper, we introduce a new experimental paradigm, Incentivized Resume Rating, for measuring employers’ preferences over candidate characteristics.

The key advantage is ability to elicit the full distribution of employer preferences.

- Estimate value of characteristics at different levels of selectivity
- Independent randomization of many characteristics allows for analysis of conditional marginal effects

Other benefits:

- Can access employers who don’t respond to cold resumes
- Can measure multiple dimensions driving employer callbacks

We deploy IRR to investigate:

- Preferences of recruiters at elite colleges for student human capital investments
- Impact of demographic characteristics, beyond current literature