The Payoff to Attending a More Selective College

Dale and Krueger

Mid-career annual earnings (Payscale.com)

- Princeton (6) $121,000
- Stanford (8) $119,000
- MIT (11) $118,000
- Yale (12) $117,000
- Notre Dame (24) $110,000
- Cornell (33) $107,000
- Duke (44) $102,000
- Wake Forest (77) $95,300
- Purdue (133) $89,000
- Indiana (378) $76,700
- Valparaiso (395) $76,200
- WVU (409) $75,800
- IUSB (975) $53,100
- Shaw (1016) $41,900

Three steps in admission process

- Students decide where to apply
- School decides whom to accept
- Given acceptances, students decides where to attend

What enters in the school’s decision

- Characteristics observed by researcher
  - SAT/GPA/AP classes/clubs

- Characteristics unobserved by researcher
  - Motivation, maturity, ambition, special skills
  - Revealed in letters of recommendation, personal statement
• Let: $x_{1i}$ be measurable characteristics
  $x_{2i}$ be unmeasured characteristics
  $w_i$ be wages
  $Q_i$ be the measure of school quality
  (like SAT)

• Model we would like to estimate
  $\ln(w_i) = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + Q_i\beta_3 + \epsilon_i$

Problem

• Can find lots of data sets with $x_1$ and $Q$
  – Can measure SAT, GPA and school quality
• Few if any will have $x_2$.
• When trying to estimate the impact of schools on outcomes, will have a major omitted variables bias

• Model we end up estimating
  $\ln(w_i) = \beta_0 + x_{1i}\beta_1 + Q_i\beta_3 + \epsilon_i$

College and Beyond (C&B)

• 23,573 Students that graduated from 34 college in 1951/76/89
  – Data from institutional/college board records
  – Survey conducted in 1995-1997 that includes
    • What schools applied & accepted
    • Annual earnings in 1995
• Final sample
  – 1976 cohort
  – Exclude HBCU
  – Include full time workers

• Does the realization of $\epsilon_i$ convey information about $Q_i$?
• Suppose that the skills schools find attractive (drive, ambition, enthusiasm) are the same things that are rewarded in the job market
• What is the bias in the coefficient on $\beta_3$ in the traditional model?
Some schools in the sample

- Public
  - Penn State, Miami (Ohio), Michigan, UNC
- Liberal arts
  - Oberlin, Kenyon, Denison
- Exclusive liberal arts
  - Swarthmore, Williams, Wellesley
- Top 20
  - Stanford, Penn, Northwestern, Duke, Georgetown

Students were asked
- Where they applied?
- Where were they accepted?

This allows the authors to group students based on where they applied/admitted

Too many possible combinations – so group into equivalence classes based on SAT
- Same “school” if average SAT in the same 25 point range

Controlling for unobservables

- Consider students that applied and were accepted to the same two schools (A&B)
  - One went to A – the other went to B
- Schools view these students as somewhat equivalent along unobserved and observed dimensions
- What key assumption does the author have to make about why one went to A and the other went to B?

Students A&B K&L applied and were accepted to same set of schools
Students F&G only applied to one school – were excluded
Student O applied to a unique set of schools and had no match
Model

- Construct dummy variable for each “group”
- Add all but one to the model
- These dummy variables capture the fact that some students are observationally similar
  - “hold constant” the characteristics that lead one to apply/get accepted at a group of schools

New model

\[ \ln(\text{earnings}_i) = \beta_0 + x_1 \beta_1 + Q_1 \beta_3 + \sum_j D(j)_i \alpha_j + \epsilon_i \]

- Let \( D(j)_i \) be a dummy variable that equals 1 if person \( i \) belongs to group \( j \)
- \( \alpha_j \) represents the relative earnings for the group compared to the omitted category

Some facts

- 70% listed another school they applied to other than the one they attended
- 62% attended the most selective school to which they were admitted
- 44% had at least one other student to which they were matched
- Final sample: 14,238
- 1,233 different applicant groups
### TABLE II
**Means and Standard Deviations of the C&B Data Set**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unweighted</th>
<th>Weighted*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full sample</td>
<td>Full sample</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>Log(earnings)</td>
<td>11.121</td>
<td>0.787</td>
</tr>
<tr>
<td>Annual earnings</td>
<td>96,768</td>
<td>62,504</td>
</tr>
<tr>
<td>(1995 dollars)</td>
<td>0.391</td>
<td>0.488</td>
</tr>
<tr>
<td>Female</td>
<td>0.059</td>
<td>0.235</td>
</tr>
<tr>
<td>Black</td>
<td>0.016</td>
<td>0.124</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.027</td>
<td>0.162</td>
</tr>
<tr>
<td>Asian</td>
<td>0.003</td>
<td>0.059</td>
</tr>
<tr>
<td>Other race</td>
<td>0.040</td>
<td>0.201</td>
</tr>
<tr>
<td>Own SAT/100</td>
<td>11.820</td>
<td>1.601</td>
</tr>
<tr>
<td>School average SAT/100</td>
<td>11.949</td>
<td>0.828</td>
</tr>
</tbody>
</table>

### TABLE III
**Log Earnings Regressions Using College and Beyond Survey, Sample of Male and Female Full-Time Workers**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full sample</th>
<th>Restricted sample</th>
<th>Similar school SAT match* model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basic model: no selection controls</td>
<td>Matched applicant model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>School-average SAT</td>
<td>0.078</td>
<td>0.046</td>
<td>0.016</td>
</tr>
<tr>
<td>Own SAT score/300</td>
<td>0.018</td>
<td>0.006</td>
<td>0.016</td>
</tr>
<tr>
<td>Female</td>
<td>0.003</td>
<td>0.019</td>
<td>0.003</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.127</td>
<td>0.110</td>
<td>0.112</td>
</tr>
<tr>
<td>N</td>
<td>14,258</td>
<td>6,335</td>
<td>6,335</td>
</tr>
</tbody>
</table>