Preferences and the gender earnings gap: Evidence on higher education choice

Aparajita Dasgupta    Anisha Sharma

Ashoka University

December 21, 2018
Can heterogeneous preferences over education and jobs explain the gender earnings gap?
Motivation

Recent evidence suggests wage gaps across subjects is as large as across school vs college education (Altonji 2014)
In the US, women are two-thirds as likely as men to study STEM (Gemici and Wiswall 2014).

In India, women are a third as likely as men to study computer science and engineering and two thirds as likely to study business (MHRD 2018).

In Indian schools, girls are 3/4 as likely as boys to study science and almost equally likely to study business (DISE 2018)
In the US, women are two-thirds as likely as men to study STEM (Gemici and Wiswall 2014).

In India, women are a third as likely as men to study computer science and engineering and two thirds as likely to study business (MHRD 2018).

In Indian schools, girls are 3/4 as likely as boys to study science and almost equally likely to study business (DISE 2018)

Is subject choice linked to the gender wage gap? Why does subject choice vary by gender?
What does this project do

- Estimate the economic return to different subject streams
  - Across STEM, business and the humanities
What does this project do

- Estimate the economic return to different subject streams
  - Across STEM, business and the humanities
- Identify determinants of subject choice
  - Incorporate measures of ability, non-cognitive traits, preferences, expectations
What does this project do

- Estimate the economic return to different subject streams
  - Across STEM, business and the humanities
- Identify determinants of subject choice
  - Incorporate measures of ability, non-cognitive traits, preferences, expectations
- Estimate a model of subject choice given preferences over subjects and occupations
Related literature

- Estimating a model of subject choice by gender:
  - Preferences rather than abilities: Gemici and Wiswall (2014)
  - Preferences for job-related attributes: Zafar (2013), Wiswall and Zafar (2017)
  - Behavioural differences: Reuben et al. (2013)
- Non-cognitive skills and education choice: Buser et al. (2014)
- Stated preference data and education choice (Weinberger 2004)
1. Introduction
2. Data
3. Returns to majors
4. Major choice
5. Conclusion
Institutional background

Ashoka University’s Young India Fellowship programme
Ashoka University’s Young India Fellowship programme

- Multiple cohorts (2011-2018) of comparable students in terms of abilities
- Distribution across subjects is similar to national distribution
- Diverse in terms of subject and socioeconomic background
- 73% on financial aid
Aims of the programme

- Identify and develop “leadership potential” among elite students
- Study a diverse range of subjects across 8 terms in 48 weeks
- Undertake a 6-month internship project in partnership with a client from the private, public or social sector
- Receive mentoring from business and academic leaders to develop their professional skills and networks
Administrative data

- Data collected as part of online application
- Academic background (class 10, 12, UG, PG)
- Socioeconomic background (location of upbringing, parent’s education and employment status, household income)
Survey data

**Subject choice module**

- Directly collected information on preferences
Survey data

Subject choice module

- Directly collected information on preferences

Labour market module

- Labour market histories,
- Expectations of future labour outcomes
Survey data

Subject choice module
- Directly collected information on preferences

Labour market module
- Labour market histories,
- Expectations of future labour outcomes

Personality traits module
- Measures of attitudes towards risk and competition, and self-esteem
Sample

We collect information on 675 jobs of 326 survey respondents across a number of occupations.

- Between 22-35 yrs old
- 54% women
- 50% science, 20% econ/business, 30% humanities majors
- Information includes wages, occupational sector, duration of work, location
We collect information on 675 jobs of 326 survey respondents across a number of occupations.

- Between 22-35 yrs old
- 54% women
- 50% science, 20% econ/business, 30% humanities majors
- Information includes wages, occupational sector, duration of work, location

Comparison with all-India data:

- Women account for 47% of enrolment.
- All India: 56% of all students enrol in arts/business in college, 41% in STEM (MHRD 2018)
Kernel density estimates

Density

Log monthly earnings

- STEM
- Econ/Business
- Humanities
Share of women among graduates

- Management
- Business and finance
- Architecture and engineering
- Natural and social sciences
- Office and administrative
- Computer and mathematical
- Sales
- Government
- Community and social
- Others
- Education
- Legal
- Media

Share of women

Women
Men

Share of women among graduates
Earnings

\[ y_{imjt} = \beta_m M_i + \beta_g \text{Male}_i + \gamma X_i + \psi_j + \delta_t + \epsilon_{imjt} \]

- \( y_{imt} \) are the earnings of person \( i \) in occupation \( j \) at time \( t \), after studying major \( m \)
- \( M_i \) is the major studied by person \( i \)
- \( X_i \) contains measures of ability, non-cognitive traits and socio-economic background
- Standard errors clustered by individual
### Earnings

<table>
<thead>
<tr>
<th>Dependent variable: log income in '000s of Rs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Econ &amp; Business</td>
<td>0.092</td>
<td>0.100</td>
<td>0.055</td>
<td>0.055</td>
<td>0.120</td>
</tr>
<tr>
<td>Arts &amp; soc sci</td>
<td>−0.225***</td>
<td>−0.162*</td>
<td>−0.202**</td>
<td>−0.222**</td>
<td>0.021</td>
</tr>
<tr>
<td>Male</td>
<td>0.148**</td>
<td>0.191**</td>
<td>0.208***</td>
<td>0.222***</td>
<td>0.152**</td>
</tr>
<tr>
<td>Work Exp (yrs)</td>
<td>0.134***</td>
<td>0.157***</td>
<td>0.124***</td>
<td>0.127***</td>
<td>0.120***</td>
</tr>
<tr>
<td>Grade 10 testscore</td>
<td>0.193***</td>
<td>0.179**</td>
<td>0.175**</td>
<td>0.107</td>
<td>0.077</td>
</tr>
<tr>
<td>English fluency</td>
<td>0.225**</td>
<td>0.110</td>
<td>0.106</td>
<td>0.121</td>
<td>0.076</td>
</tr>
<tr>
<td>Risk measure</td>
<td>0.002</td>
<td>0.002</td>
<td>−0.005</td>
<td>(0.012)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Competitiveness</td>
<td>−0.030</td>
<td>−0.034</td>
<td>−0.004</td>
<td>(0.046)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>0.013</td>
<td>0.015*</td>
<td>0.010</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Admissions score</td>
<td>0.086**</td>
<td>0.083**</td>
<td>0.085**</td>
<td>(0.038)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Year effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HH controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Occupation controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>675</td>
<td>670</td>
<td>611</td>
<td>611</td>
<td>611</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.27</td>
<td>0.30</td>
<td>0.32</td>
<td>0.32</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Explaining major choice

- Ability
- Non-cognitive measures (attitudes towards risk, competition, self-esteem)
- Preferences
- Expectations
Explaining major choice

- Ability
- Non-cognitive measures (attitudes towards risk, competition, self-esteem)
- Preferences
- Expectations
Distribution of ability by gender

- Women
- Men

Grade 10 math scores

Density

Data

Returns to majors

Major choice

Conclusion
### Explaining major choice

- Ability
- **Non-cognitive measures** (attitudes towards risk, competition, self-esteem)
- Preferences
- Expectations
Explaining major choice

- Ability
- Non-cognitive measures (attitudes towards risk, competition, self-esteem)
- Preferences
- Expectations
Subject choice model

Student $i$ faces a choice across $M$ different majors derives utility from choosing the $m$th major of

$$U_{im} = u(Y_{im}, Z_{im}, X_i) + \epsilon_{im}$$

- $Y_{1im}, Y_{2im}, \ldots Y_{Kim}$ are major-related characteristics
- $Z_{1im}, Z_{2im}, \ldots Z_{Jim}$ are job-related characteristics
- $X_{1i}, X_{2i}, \ldots Y_{Li}$ are individual characteristics
- $\epsilon_{im}$ is any unobserved preference of student $i$ for major $m$
Subject choice model

- **Major-related characteristics**: love for the subject, expected grades, training for a specific career
- **Job-related characteristics**: expected wages, work-life balance
- **Individual characteristics**: desire to please parents, peer pressure, influence of teachers and mentors

A student will select major $m$ over major $m'$ if $EU_{im} > EU_{im'}$ for all $m \neq m'$. 
Approach in the literature so far:

- Make assumptions about expectations and use observed choices to estimate model parameters
- Elicit expectations and estimate model parameters
- Collect stated preferences
How important were each of the following factors in your decision to study [chosen major]?

- To learn more about subjects I enjoyed studying or was interested in
- To get high grades on the coursework
- ...etc...

Respondents choose from a scale of 0 to 3 where 0 is “not at all important” and 3 is “extremely important”.
Do stated preferences vary by gender?
Preferences by gender

Use an ordered probit to identify predictors of how important each factor is:

- Gender
- Ability
- Personality
- Socioeconomic background
### Preferences

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Love for subject</th>
<th>Grades</th>
<th>Wages</th>
<th>Career</th>
<th>Work-life balance</th>
<th>Parents</th>
<th>Peers</th>
<th>Teachers</th>
<th>Option value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.613***</td>
<td>-0.450***</td>
<td>0.101</td>
<td>-0.298**</td>
<td>-0.130</td>
<td>0.225*</td>
<td>0.380***</td>
<td>0.103</td>
<td>-0.210*</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.128)</td>
<td>(0.128)</td>
<td>(0.131)</td>
<td>(0.126)</td>
<td>(0.134)</td>
<td>(0.133)</td>
<td>(0.127)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>Grade 10 score</td>
<td>-0.032***</td>
<td>-0.001</td>
<td>0.033***</td>
<td>0.016</td>
<td>0.004</td>
<td>0.043***</td>
<td>0.045***</td>
<td>0.004</td>
<td>0.034***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Risk measure</td>
<td>-0.031</td>
<td>-0.008</td>
<td>0.022</td>
<td>0.005</td>
<td>0.024</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.021</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Competitiveness</td>
<td>0.223***</td>
<td>0.291***</td>
<td>0.189**</td>
<td>0.237***</td>
<td>0.187*</td>
<td>0.090</td>
<td>0.076</td>
<td>0.309***</td>
<td>0.247***</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.090)</td>
<td>(0.093)</td>
<td>(0.091)</td>
<td>(0.096)</td>
<td>(0.088)</td>
<td>(0.096)</td>
<td>(0.088)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>0.019</td>
<td>-0.014</td>
<td>-0.022</td>
<td>0.012</td>
<td>-0.027***</td>
<td>-0.025*</td>
<td>-0.017</td>
<td>-0.005</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Mother works</td>
<td>-0.286***</td>
<td>-0.251**</td>
<td>-0.299**</td>
<td>-0.333***</td>
<td>-0.168</td>
<td>0.040</td>
<td>-0.013</td>
<td>-0.026</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.123)</td>
<td>(0.127)</td>
<td>(0.125)</td>
<td>(0.128)</td>
<td>(0.128)</td>
<td>(0.131)</td>
<td>(0.130)</td>
<td>(0.131)</td>
</tr>
<tr>
<td>Mother has UG degree</td>
<td>0.230</td>
<td>0.021</td>
<td>-0.142</td>
<td>-0.163</td>
<td>-0.229</td>
<td>0.140</td>
<td>0.094</td>
<td>0.183</td>
<td>0.247</td>
</tr>
<tr>
<td></td>
<td>(0.186)</td>
<td>(0.183)</td>
<td>(0.156)</td>
<td>(0.173)</td>
<td>(0.163)</td>
<td>(0.187)</td>
<td>(0.189)</td>
<td>(0.189)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>Year effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
<td>322</td>
</tr>
<tr>
<td>Log pseudolikelihood</td>
<td>-347.36</td>
<td>-419.23</td>
<td>-422.30</td>
<td>-401.40</td>
<td>-428.23</td>
<td>-431.95</td>
<td>-378.51</td>
<td>-421.76</td>
<td>-406.99</td>
</tr>
</tbody>
</table>
Preferences vary by gender

Women were more likely to state the following factors were “very important”

- Love for the subject: 24 pp
- Expected grades: 8.5 pp
- Career path: 12.2 pp
- Option value: 8.0 pp

Men were more likely to state the following factors were “very important”

- Desire to please parents: 6.6 pp
- Peer pressure: 5.8 pp
Sensitivity of choice to parameters

Do changes in the reported preferences predict changes in choice of major?
Do changes in the reported preferences predict changes in choice of major?

Estimate a multinomial logit model of subject choice as a function of the stated preferences separately for men and women and find that different preferences are driving selection into different majors.
## University preferences

<table>
<thead>
<tr>
<th></th>
<th>Science</th>
<th>Business</th>
<th>Humanities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love of subject</td>
<td>-0.12***</td>
<td>-0.09</td>
<td>-0.11*</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Grades</td>
<td>-0.07*</td>
<td>-0.05</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Wages</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.10*</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Career</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Work-life balance</td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Parents</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Teacher</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Peers</td>
<td>0.10**</td>
<td>0.17***</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Option value</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Observations</td>
<td>326</td>
<td>150</td>
<td>176</td>
</tr>
</tbody>
</table>
Preferences vary by gender

For women, rating the following as “important” vs “unimportant” predicts:

- Love for the subject: 11 pp less likely to study science
- Expected wages: 10 pp more likely to study science

For men, rating the following as “important” vs “unimportant” predicts:

- Peer pressure: 17 pp more likely to study science
Expectations about future labour market outcomes

No systematic gender differences across
- Possibility of future employment
- Possibility of getting their dream job

Some evidence that men predict 10-20% higher incomes for themselves over 10 and 20-year periods.
Results: summary

**Gender matters**
- There are large gender gaps in wages
- These emerge from both occupational choice and major choice
  - Women select into majors associated with lower wages
  - Women in high-earnings majors select into lower paying occupations
- No evidence these are driven by differences in cognitive abilities or non-cognitive personality measures
Preferences for majors are heterogeneous by gender

- Women prioritise undergraduate benefits (love of subject, higher grades)
- Men are particularly influenced by parents, peers

Shifts in preference parameters predict shifts in major choice

- Higher weight on expected wages shifts women into STEM and economics/business
- Higher weight on interest in subject shifts women into the humanities
- Lower weight on preferences of parents and peers shifts men into the humanities
Next steps

- Elicit expectations data on outcomes associated with majors
- Elicit expectations data on outcomes associated with jobs
- Estimate models of job and major choice