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

Aparajita Dasgupta Anisha Sharma

Ashoka University

December 21, 2018

# Question

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)

# Motivation

- 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)

# Motivation

- 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

- Return to a major: Altonji et al. (2012, 2014), Gemici & Wiswall (2014), Sahoo and Klaasen (2018)
- 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)
- Subjective expectations and education choice: Dominitz and Manski (1996), Attanasio and Kaufman (2014)
- Stated preference data and education choice (Weinberger 2004)

# Institutional background

Ashoka University's Young India Fellowship programme

# Institutional background

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

# YIF

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

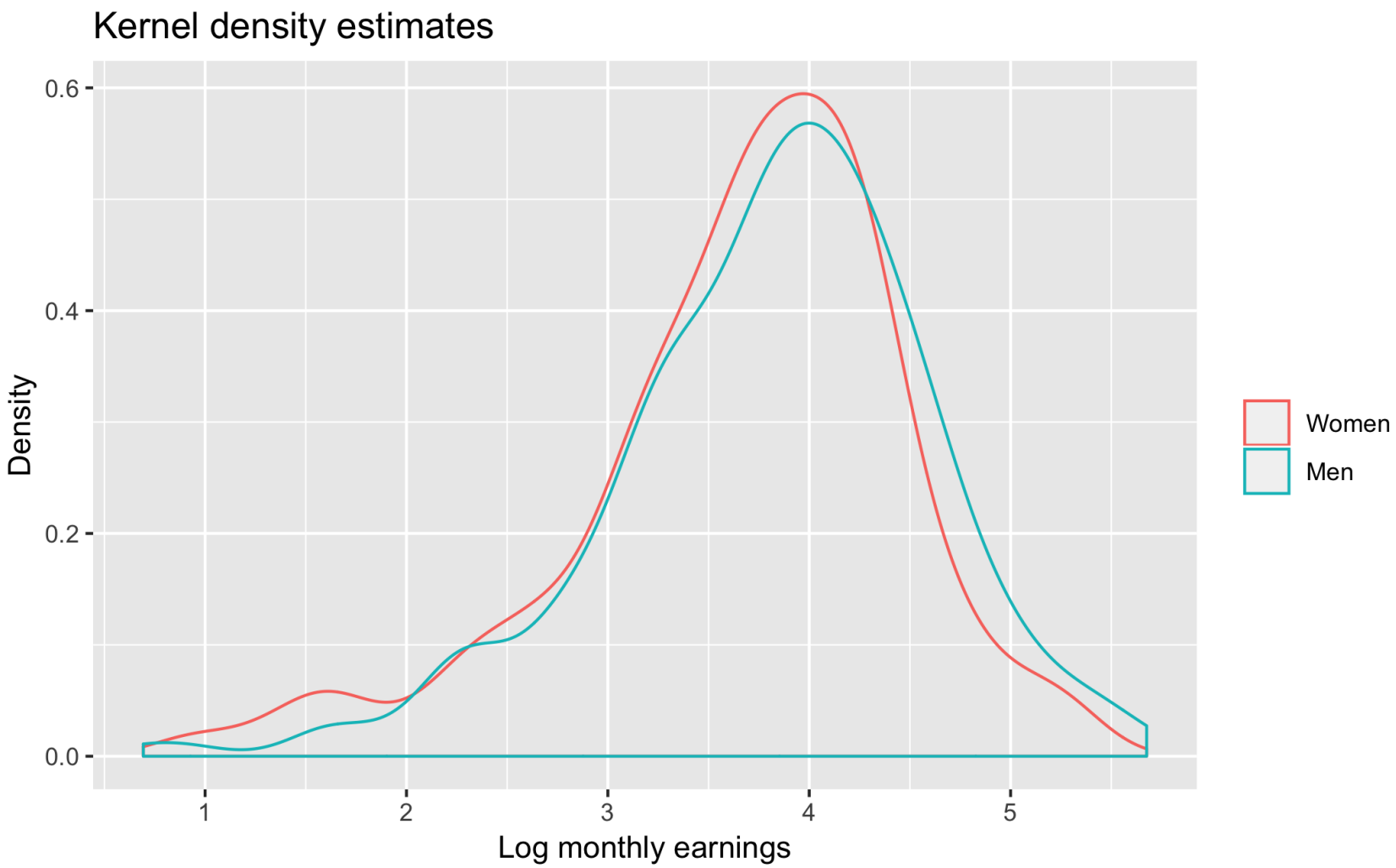
# Sample

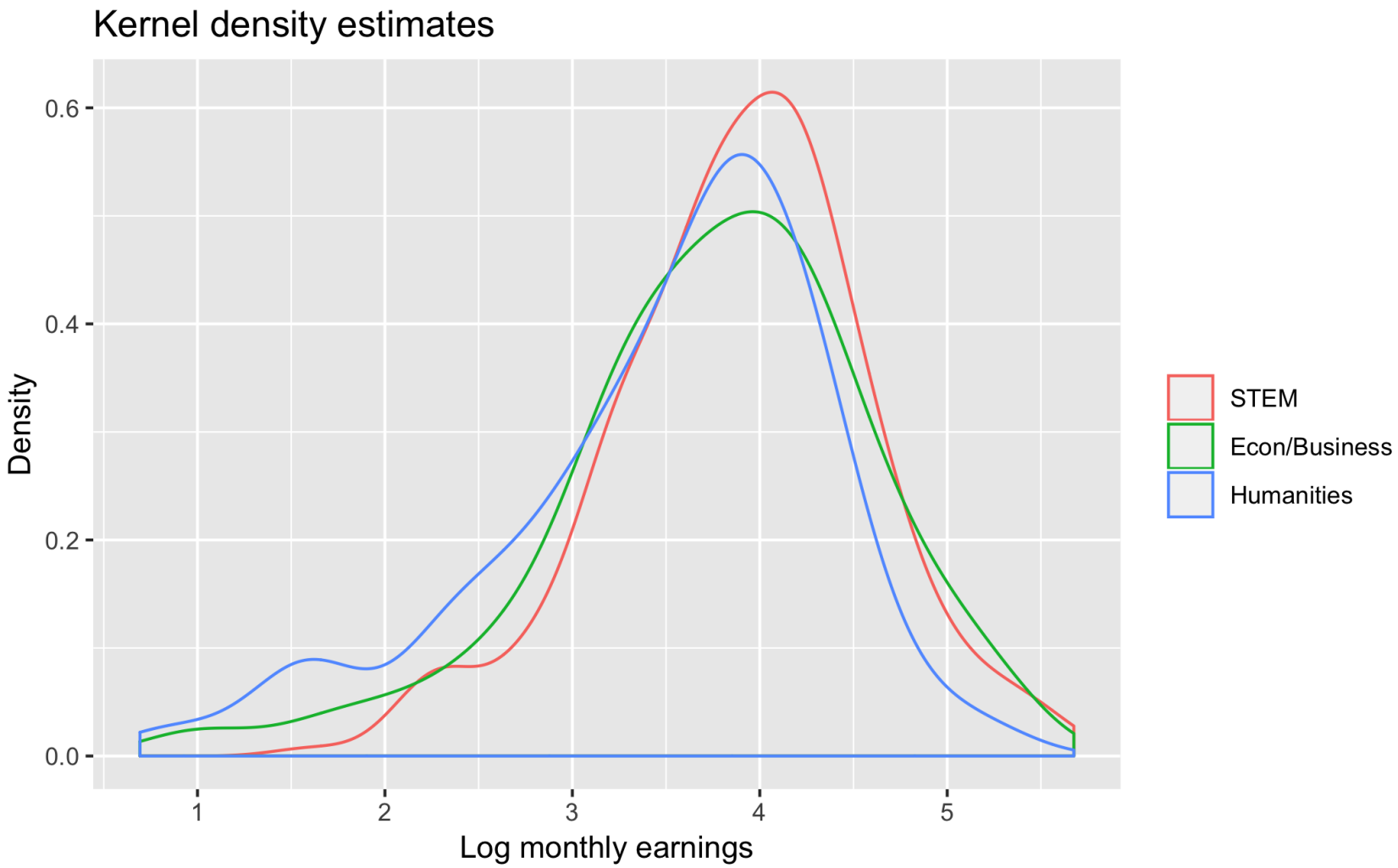
We collect information on 675 jobs of 326 survey respondents across a number of occupations.

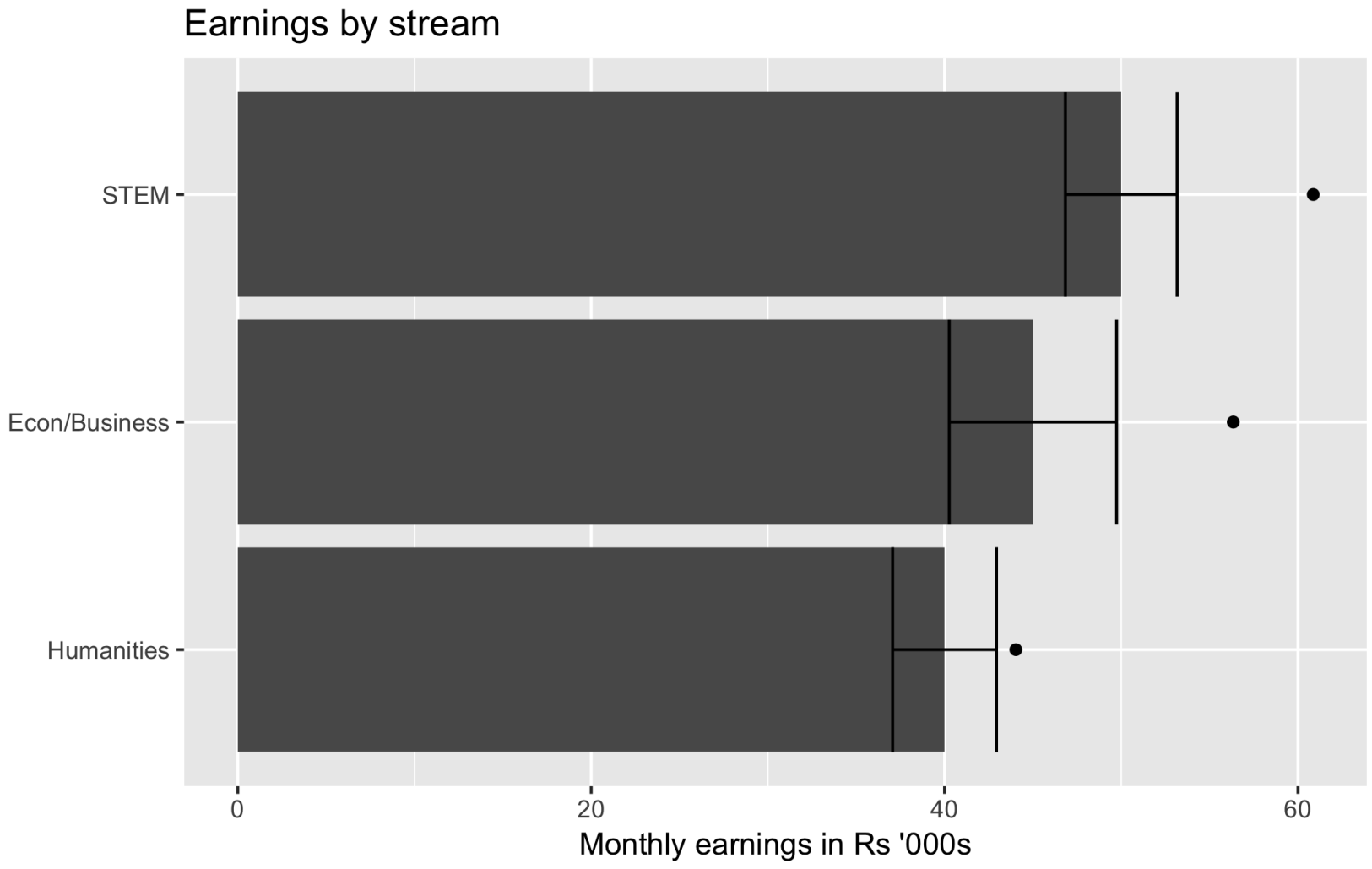
- Between 22-35 yrs old
- 54% women
- 50% science, 20% econ/business, 30% humanities majors
- Information includes wages, occupational sector, duration of work, location

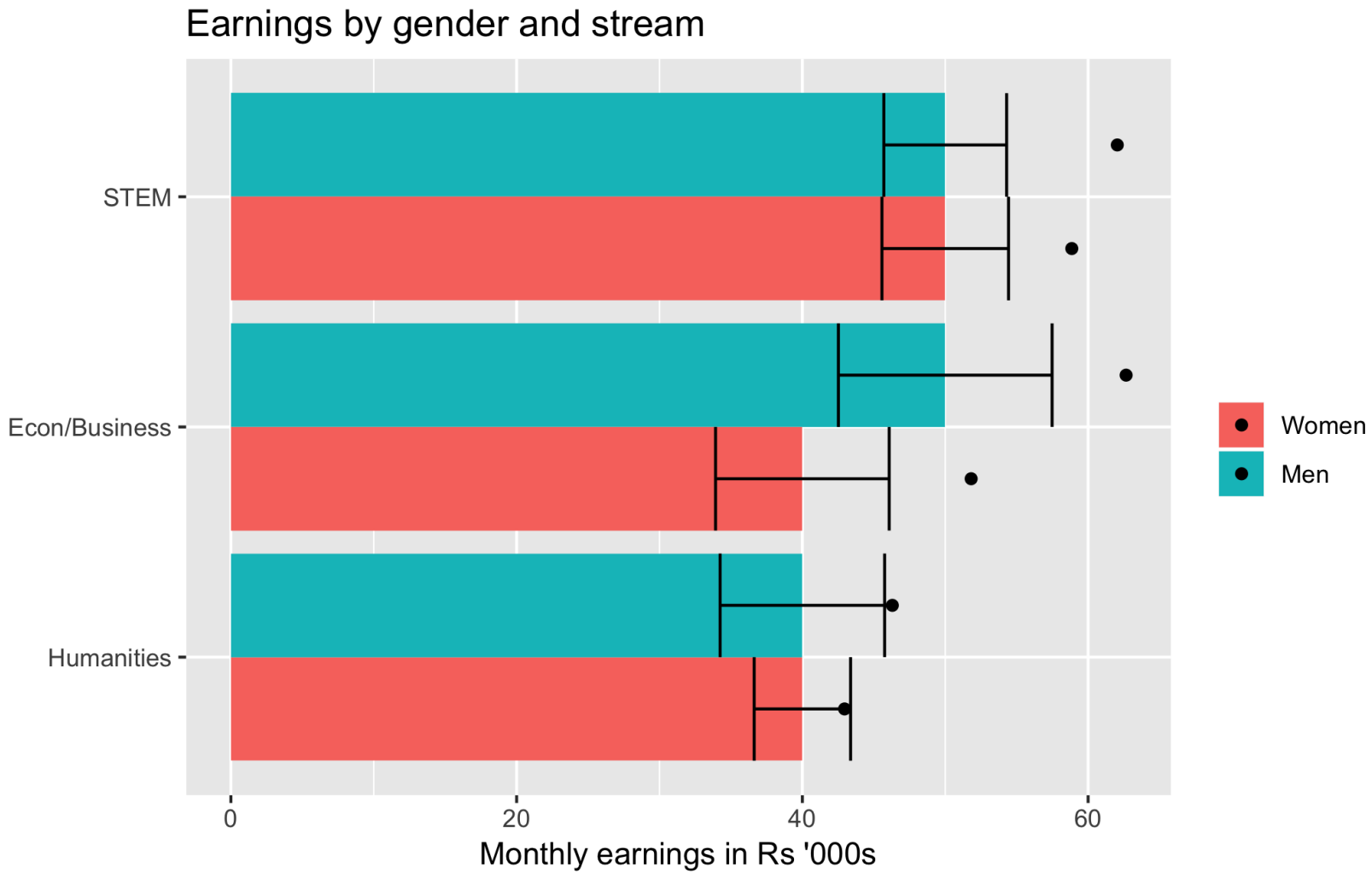
Comparison with all-India data:

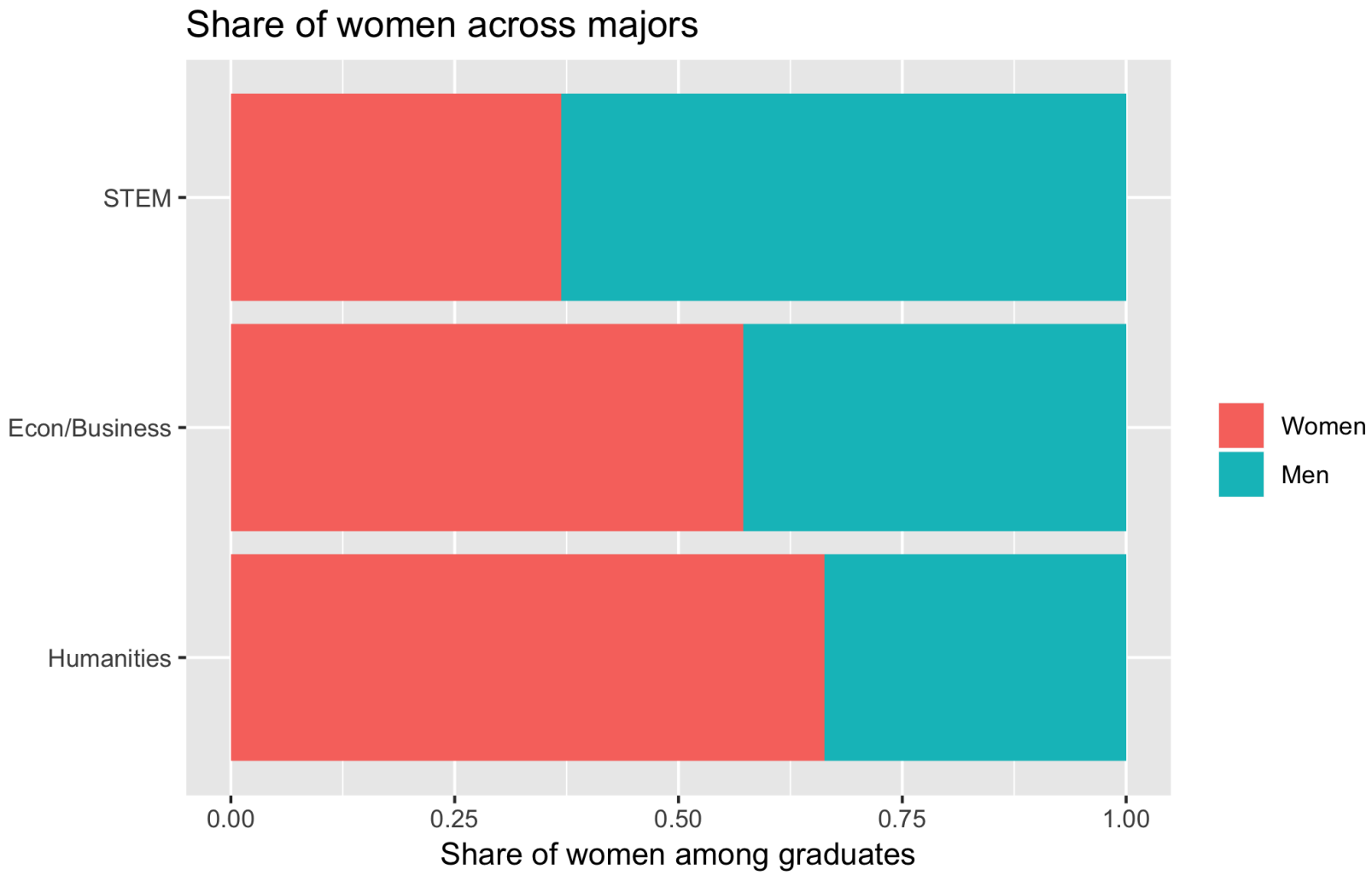
- Women account for 47% of enrolment.
- **All India**: 56% of all students enrol in arts/business in college, 41% in STEM (MHRD 2018)

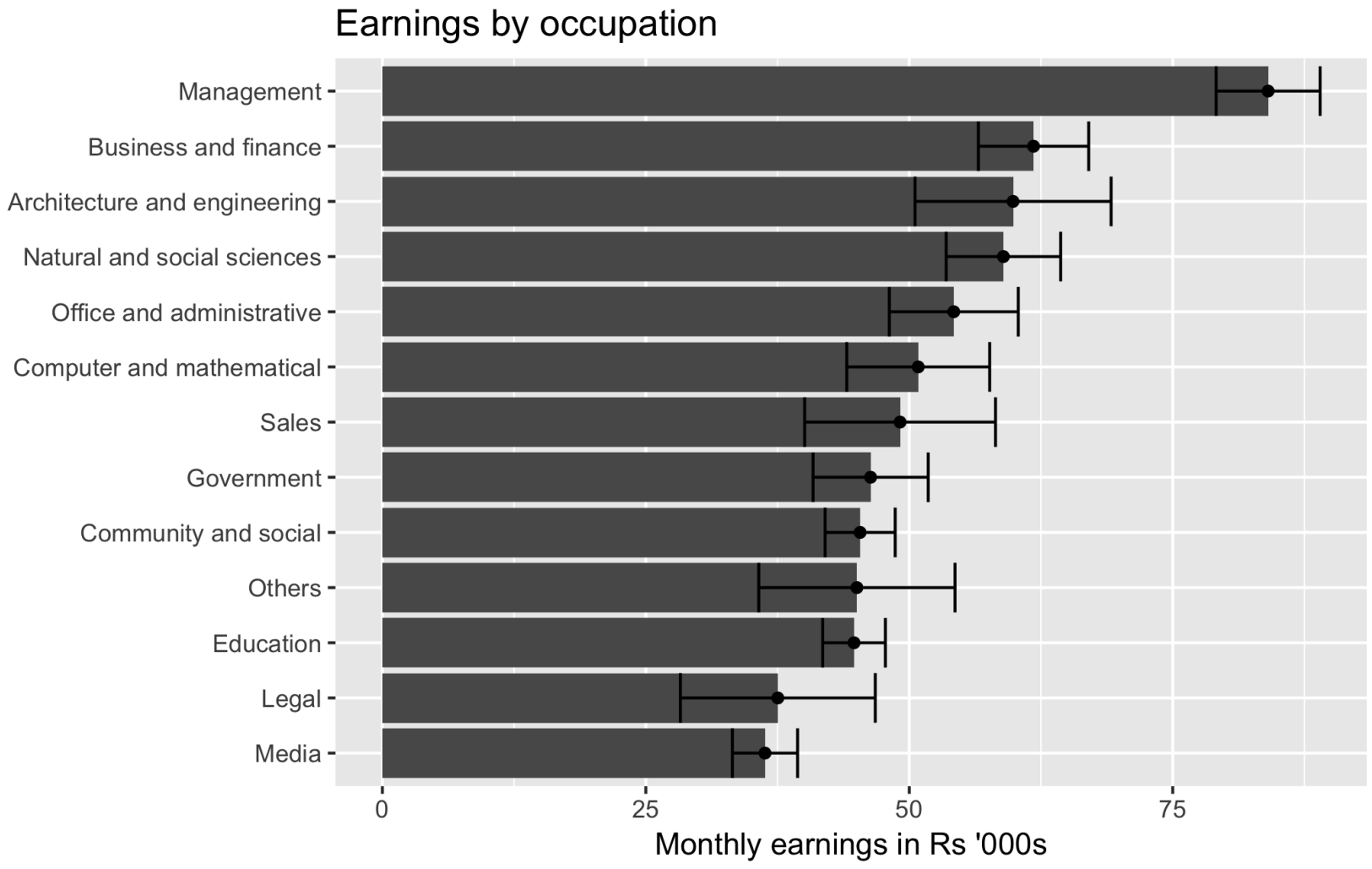

Kernel density estimates
Density
0.6
0.4
0.2
0.0
1
2
3
4
5
Log monthly earnings
Women
Men

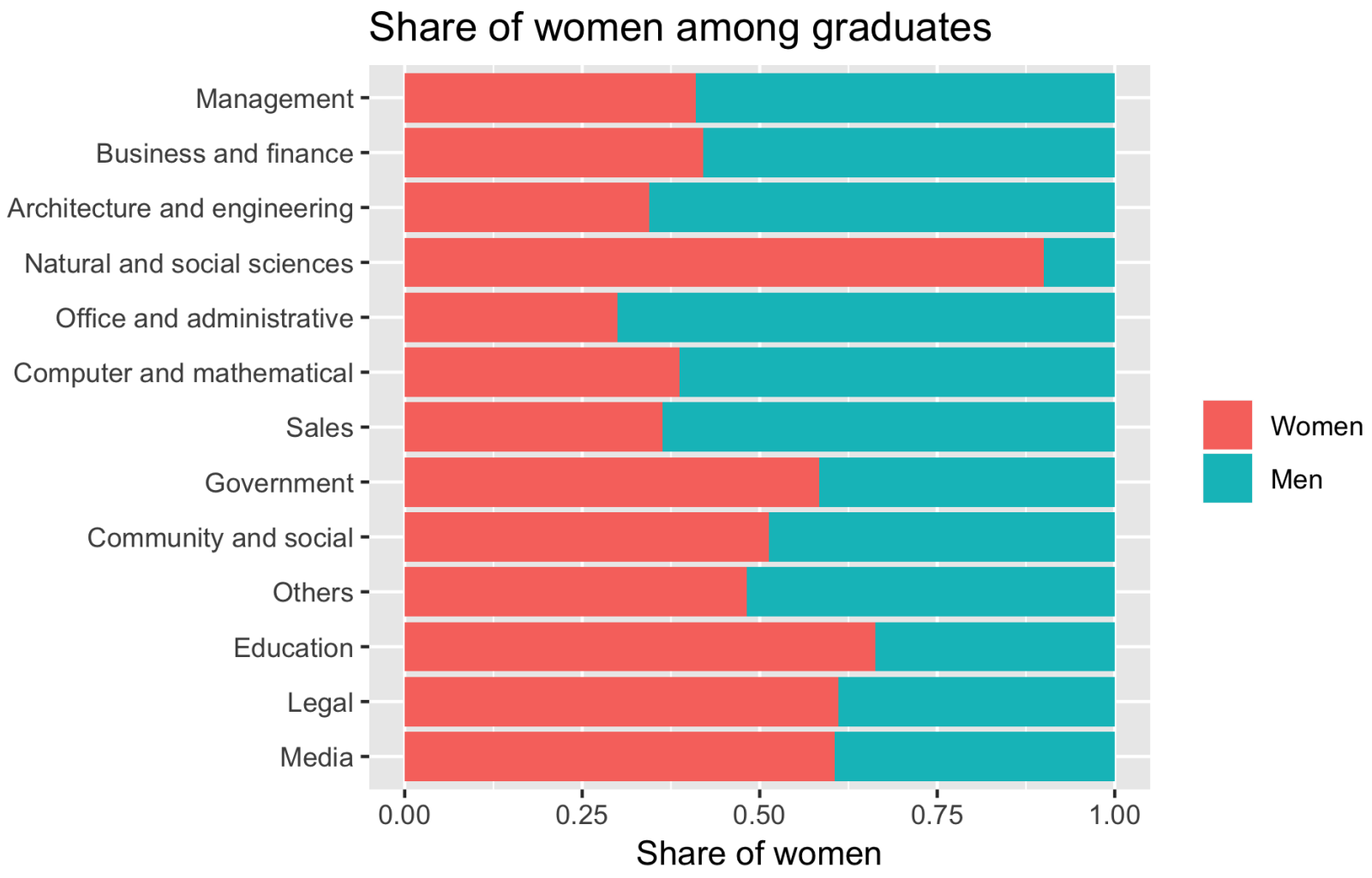

Kernel density estimates
Density
0.6
0.4
0.2
0.0
1
2
3
4
5
Log monthly earnings
STEM
Econ/Business
Humanities

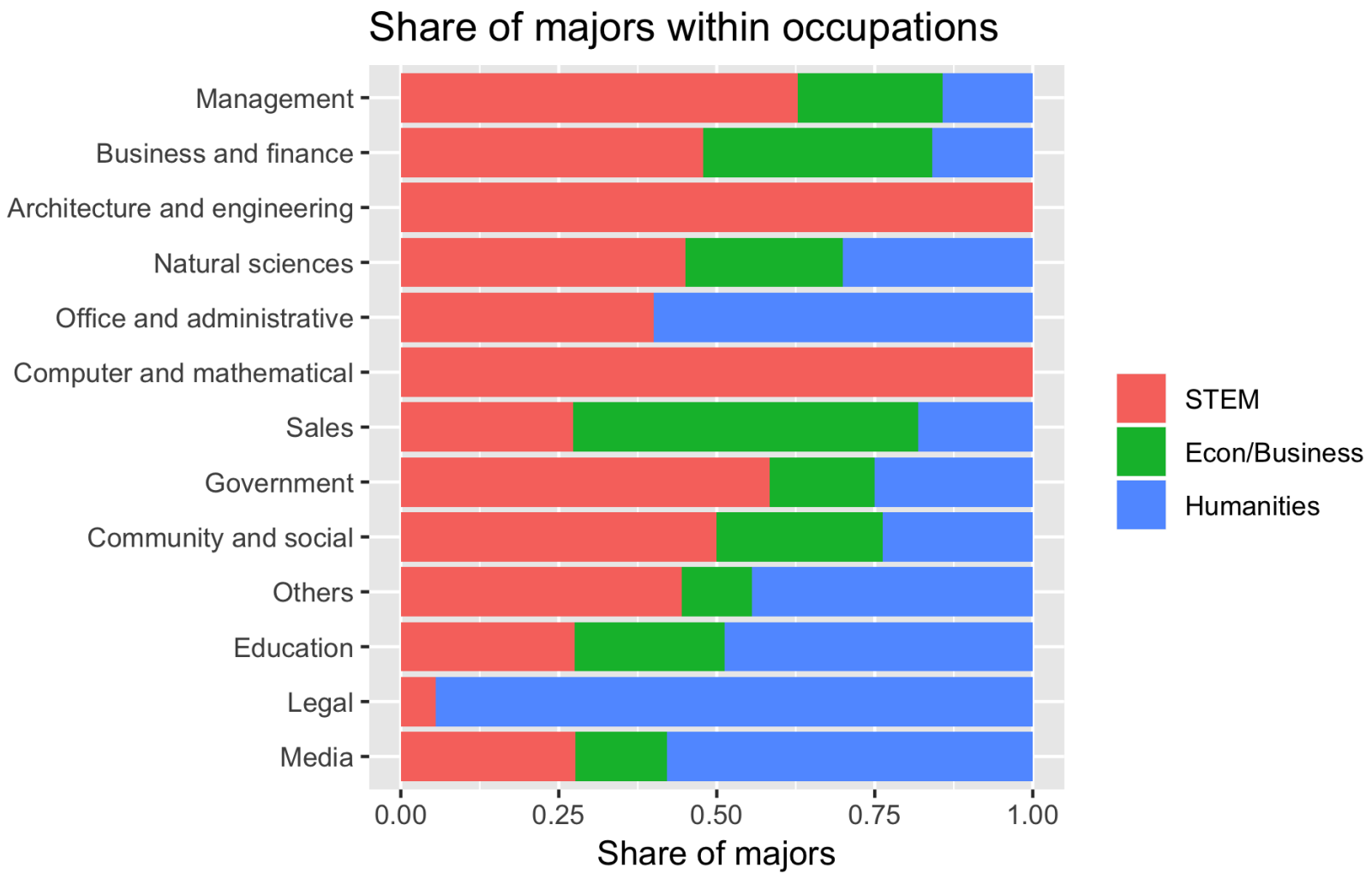

Earnings by stream
STEM
Econ/Business
Humanities
0
20
40
60
Monthly earnings in Rs '000s

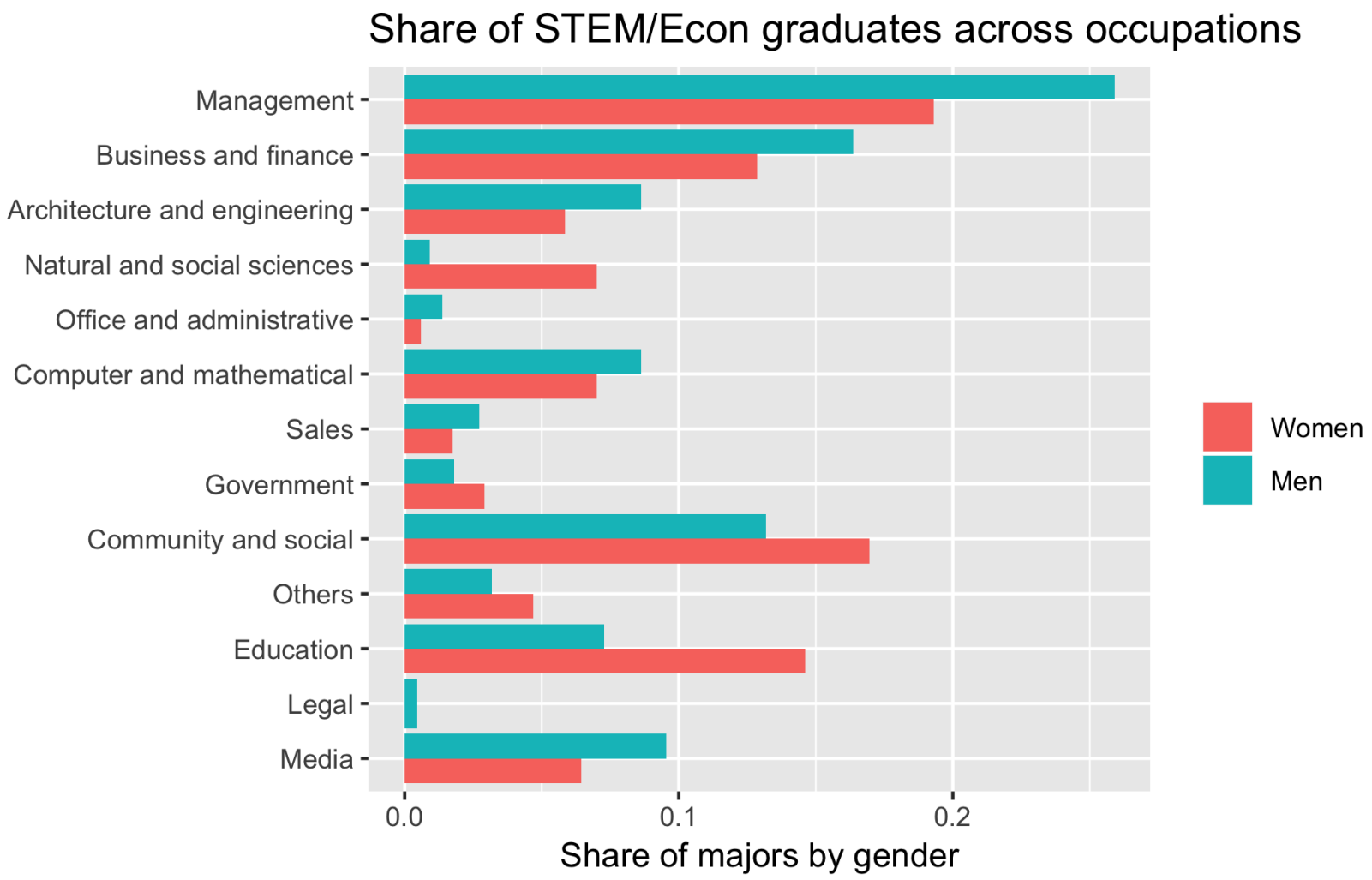

Earnings by gender and stream
STEM
Econ/Business
Humanities
0
20
40
60
Monthly earnings in Rs '000s
Women
Men


Share of women across majors
STEM
Econ/Business
Humanities
0.00
0.25
0.50
0.75
1.00
Share of women among graduates
Women
Men


Earnings by occupation
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
0
25
50
75
Monthly earnings in Rs '000s


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
0.00
0.25
0.50
0.75
1.00
Share of women
Women
Men


Share of majors within occupations
Management
Business and finance
Architecture and engineering
Natural sciences
Office and administrative
Computer and mathematical
Sales
Government
Community and social
Others
Education
Legal
Media
0.00
0.25
0.50
0.75
1.00
Share of majors
STEM
Econ/Business
Humanities


Share of STEM/Econ graduates across occupations
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
0.0
0.1
0.2
Share of majors by gender
Women
Men

# Earnings

$$y_{imjt} = \beta_m M_i + \beta_g Male_i + \gamma \mathbf{X_i} + \psi_j + \delta_t + \epsilon_{imjt}$$

- $y_{imt}$ 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 errrors clustered by individual

# Earnings

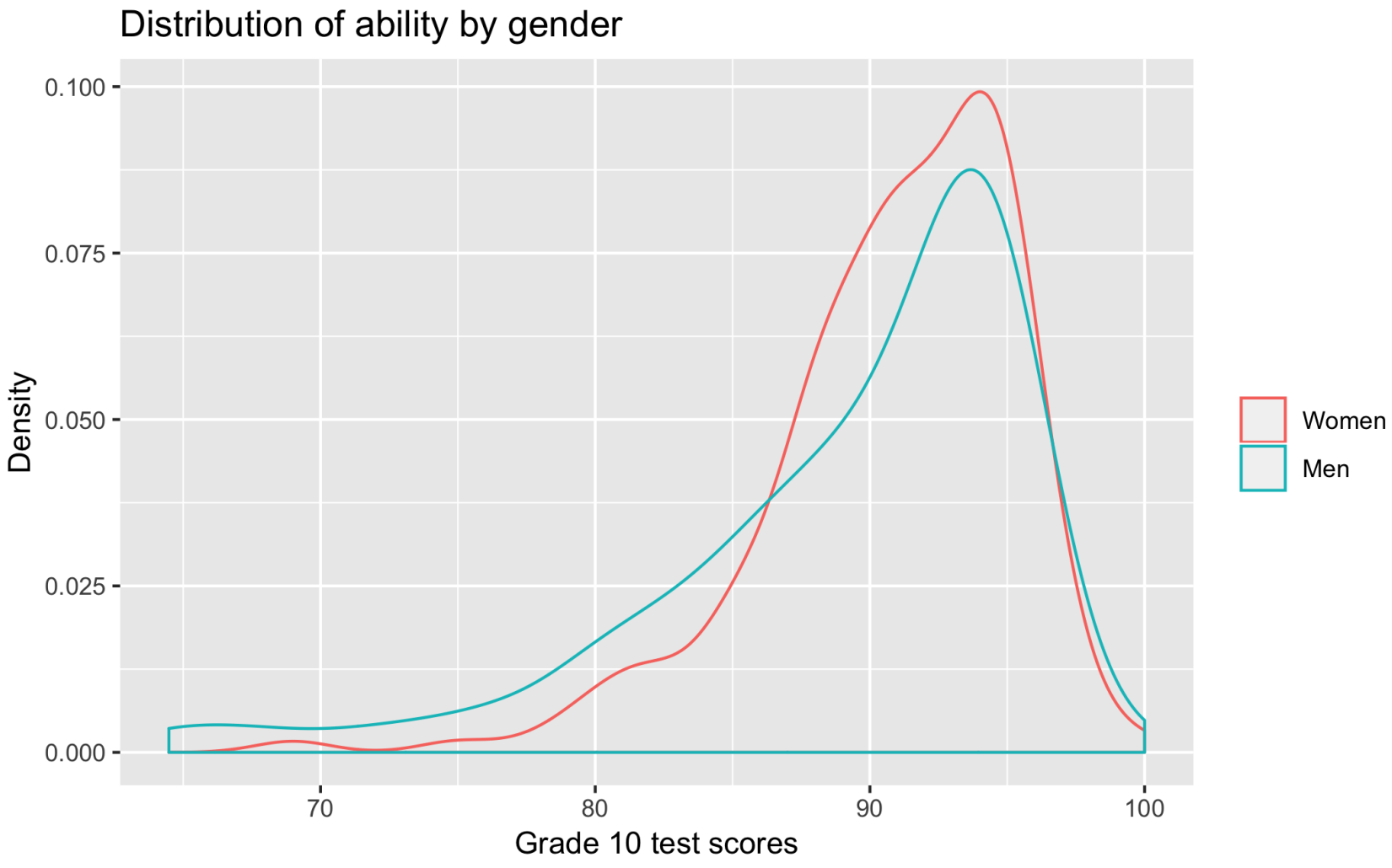
| | Dependent variable: log income in '000s of Rs | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Econ & Business | 0.092 | 0.100 | 0.055 | 0.055 | 0.120 |
| | (0.099) | (0.103) | (0.106) | (0.104) | (0.098) |
| Arts & soc sci | –0.225*** | –0.162* | –0.202** | –0.222** | 0.021 |
| | (0.084) | (0.086) | (0.093) | (0.090) | (0.092) |
| Male | 0.148** | 0.191** | 0.208*** | 0.222*** | 0.152** |
| | (0.073) | (0.075) | (0.076) | (0.075) | (0.067) |
| Work Exp (yrs) | 0.134*** | 0.157*** | 0.124*** | 0.127*** | 0.120*** |
| | (0.018) | (0.021) | (0.025) | (0.024) | (0.022) |
| Grade 10 testscore | | 0.193*** | 0.179** | 0.175** | 0.107 |
| | | (0.073) | (0.081) | (0.082) | (0.077) |
| English fluency | | 0.225** | 0.110 | 0.106 | 0.121 |
| | | (0.089) | (0.087) | (0.082) | (0.076) |
| Risk measure | | | 0.002 | 0.002 | –0.005 |
| | | | (0.012) | (0.012) | (0.011) |
| Competitiveness | | | –0.030 | –0.034 | –0.004 |
| | | | (0.046) | (0.045) | (0.044) |
| Self-esteem | | | 0.013 | 0.015* | 0.010 |
| | | | (0.008) | (0.008) | (0.008) |
| Admissions score | | | 0.086** | 0.083** | 0.085** |
| | | | (0.038) | (0.037) | (0.033) |
| Year effects | Yes | Yes | Yes | Yes | Yes |
| HH controls | No | No | No | No | Yes |
| Occupation controls | No | No | No | Yes | No |
| Observations | 675 | 670 | 611 | 611 | 611 |
| $R^2$ | 0.27 | 0.30 | 0.32 | 0.32 | 0.41 |

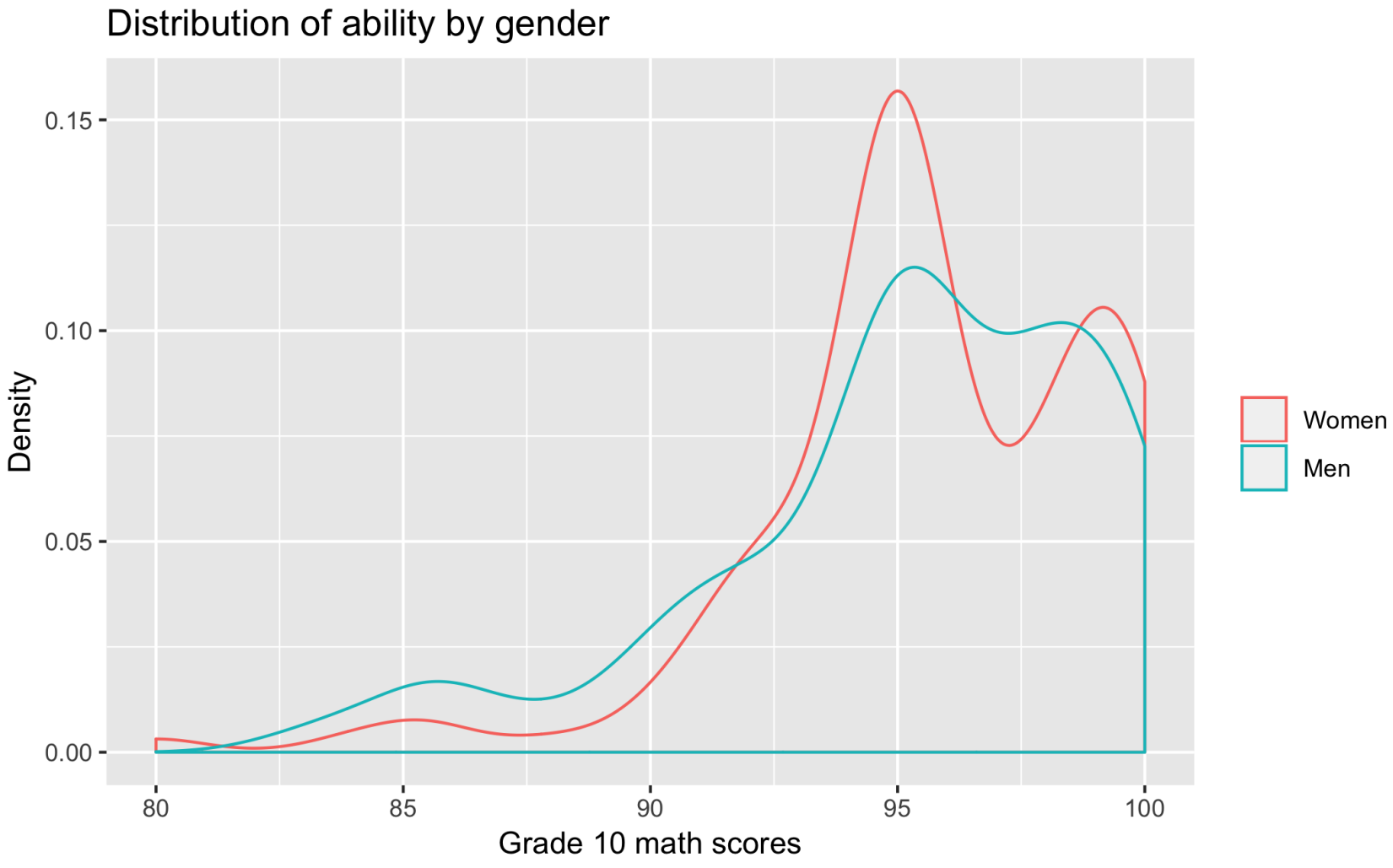
# Explaining major choice

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

# Explaining major choice

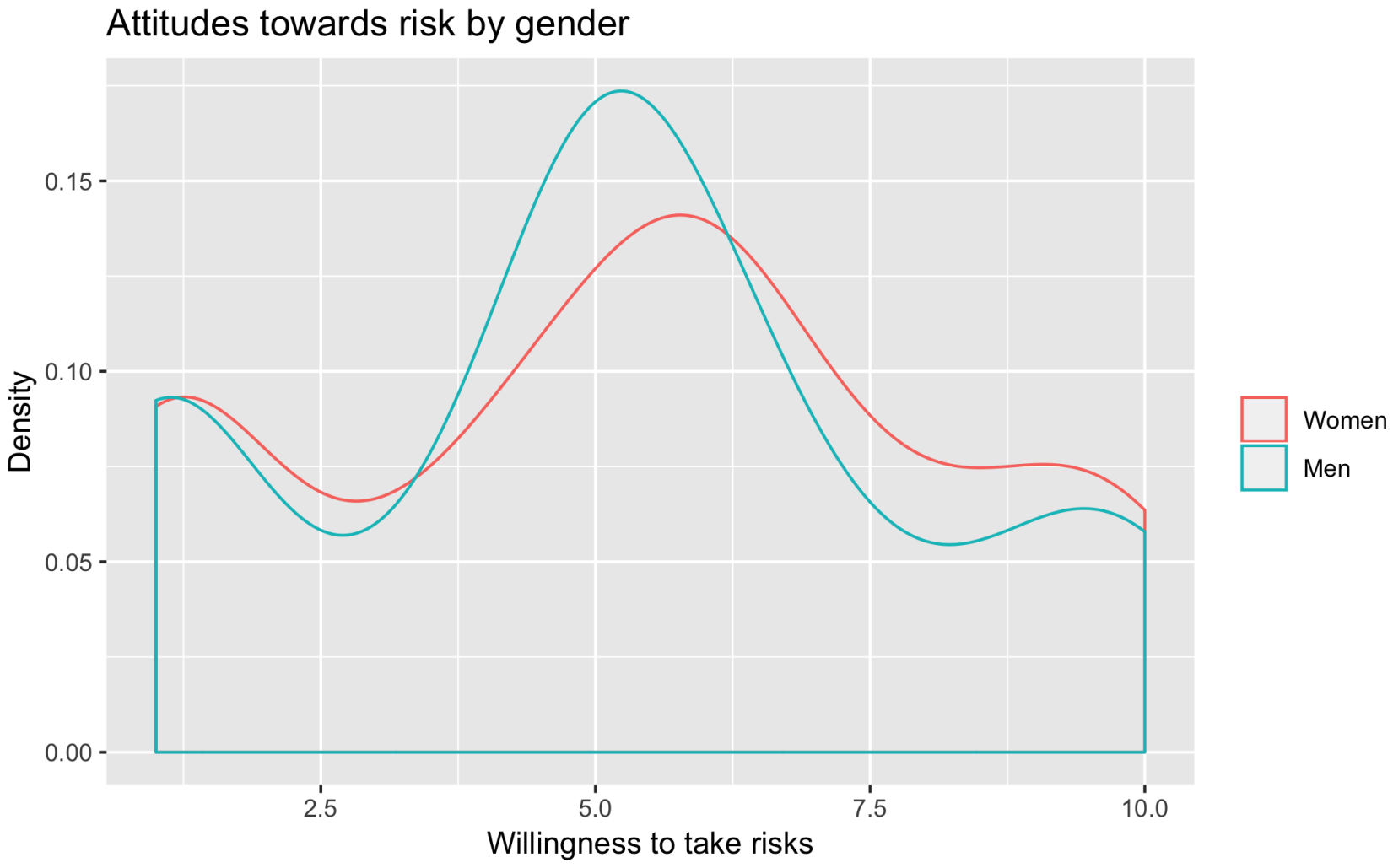
- **Ability**
- Non-cognitive measures (attitudes towards risk, competition, self-esteem)
- Preferences
- Expectations

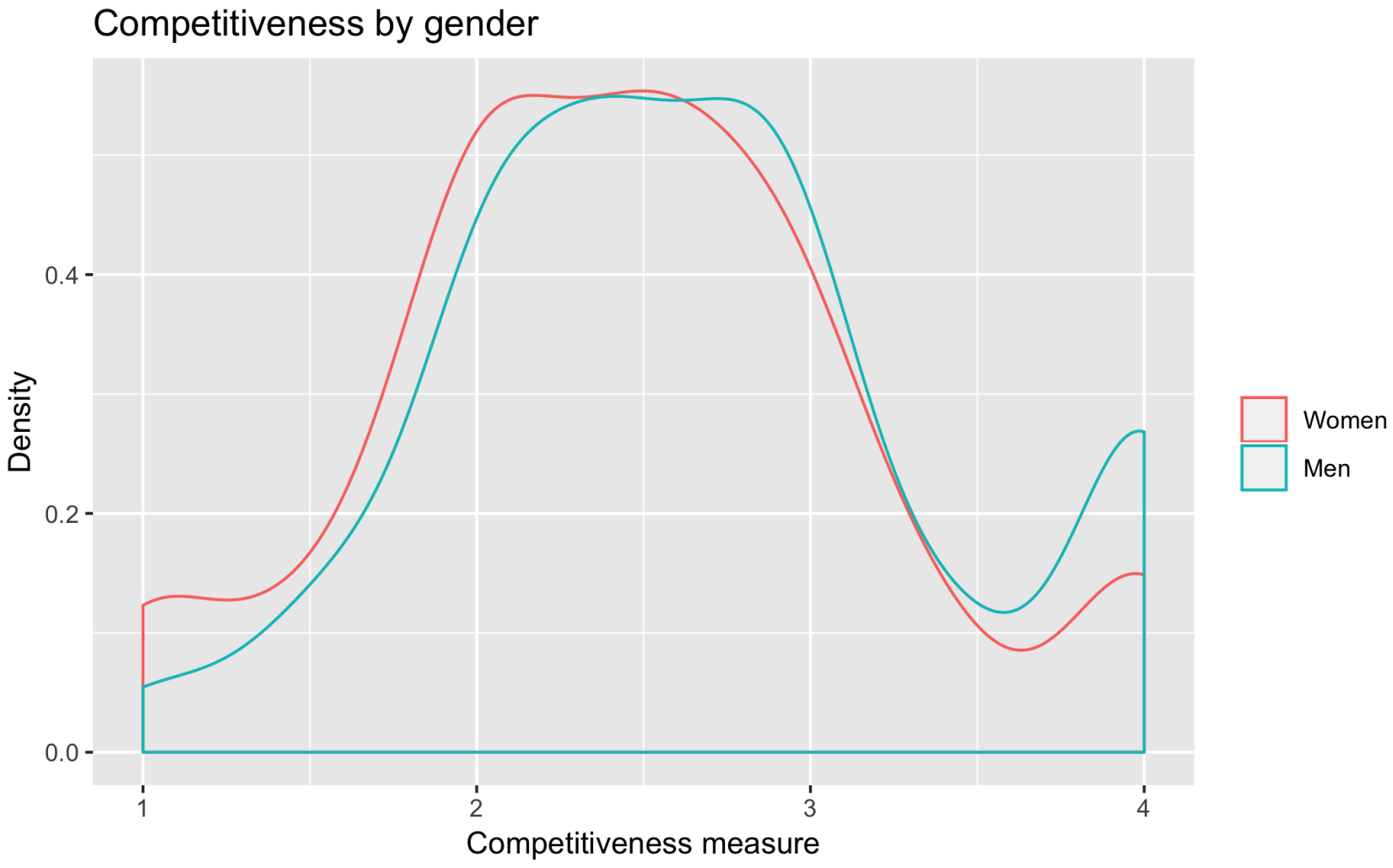

Distribution of ability by gender
0.100
0.075
0.050
0.025
0.000
Density
70
80
90
100
Grade 10 test scores
Women
Men

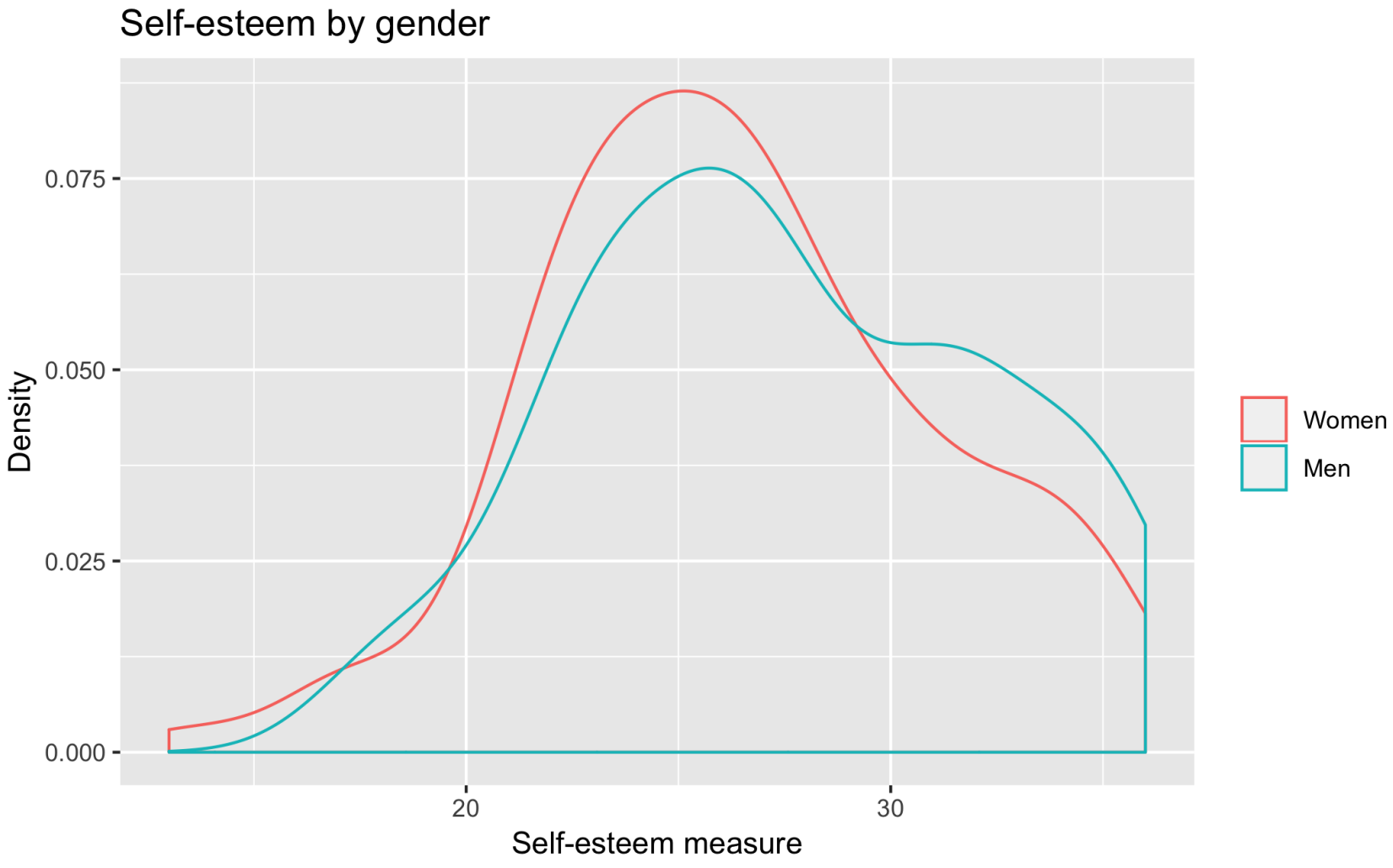

Distribution of ability by gender
Density
0.15
0.10
0.05
0.00
80
85
90
95
100
Grade 10 math scores
Women
Men

# Explaining major choice

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


Attitudes towards risk by gender
Density
0.15
0.10
0.05
0.00
2.5
5.0
7.5
10.0
Willingness to take risks
Women
Men


Competitiveness by gender
Density
0.4
0.2
0.0
1
2
3
4
Competitiveness measure
Women
Men


Self-esteem by gender
Density
0.075
0.050
0.025
0.000
20
30
Self-esteem measure
Women
Men

# 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

$$U_{im} = u(Y_{im}, Z_{im}, X_i) + \epsilon_{im} \quad (1)$$

- $Y_{1im}, Y_{2im}, ...Y_{Kim}$ are major-related characteristics
- $Z_{1im}, Z_{2im}, ...Z_{Jim}$ are job-related characteristics
- $X_{1i}, X_{2i}, ...Y_{Li}$ 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 $m \neq m'$

# Estimating a choice model

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

# Survey questions on preferences

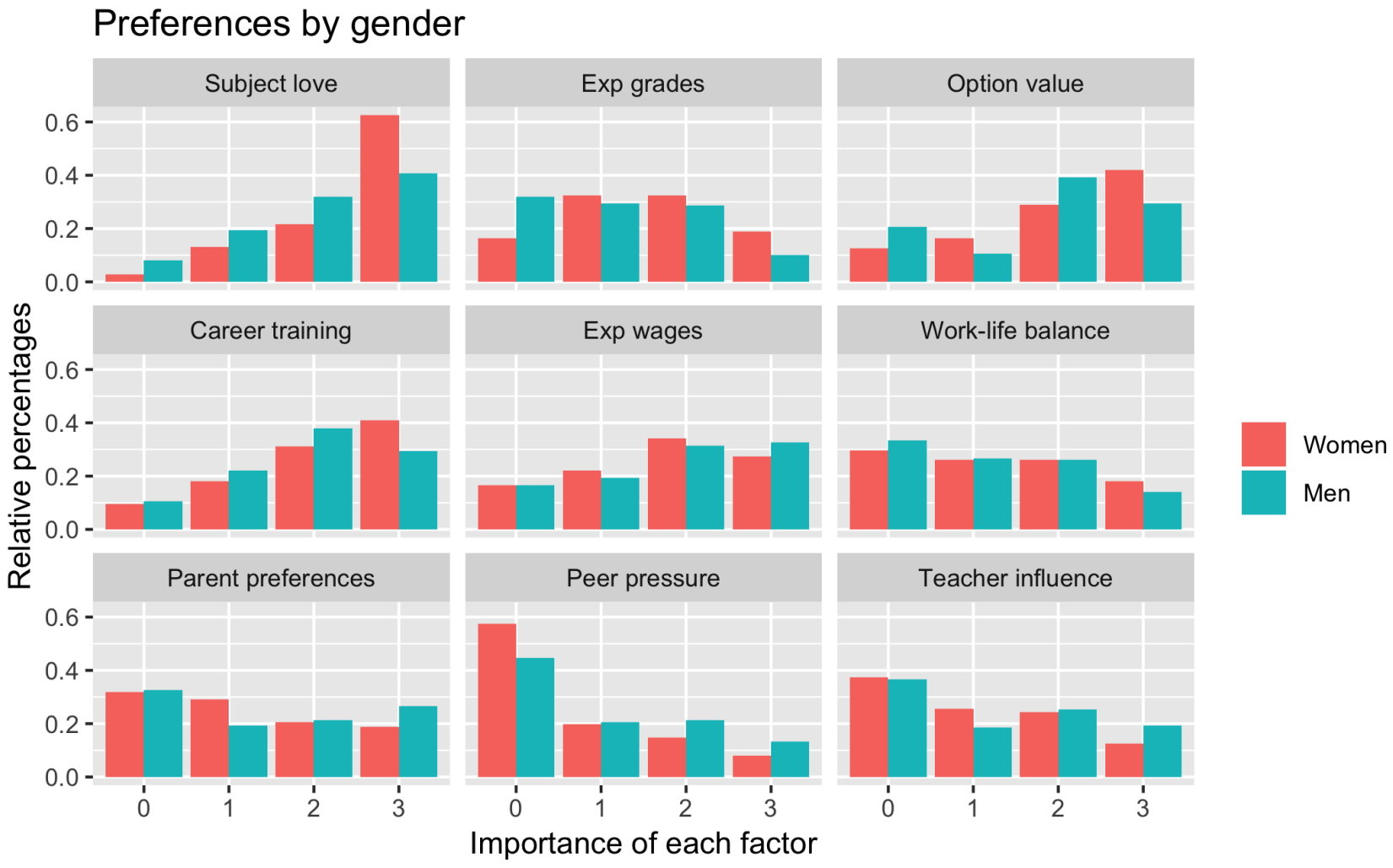
How important were each of the following factors in your decision to study [chosen major]?

- 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”.

# Preferences

Do stated preferences vary by gender?


Preferences by gender
Subject love
Exp grades
Option value
Career training
Exp wages
Work-life balance
Parent preferences
Peer pressure
Teacher influence
Relative percentages
0.0
0.2
0.4
0.6
0
1
2
3
Importance of each factor
Women
Men

# Preferences by gender

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

- Gender
- Ability
- Personality
- Socioeconomic background

# Preferences

| | Dependent variable: 0=Not important, 1=Somewhat important, 2=Moderately important, 3=Very important | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Love for subject | Grades | Wages | Career | Work-life balance | Parents | Peers | Teachers | Option value |
| Male | –0.613*** | –0.450*** | 0.101 | –0.298** | –0.130 | 0.225* | 0.380*** | 0.103 | –0.210* |
| | (0.138) | (0.128) | (0.128) | (0.131) | (0.126) | (0.134) | (0.133) | (0.127) | (0.127) |
| Grade 10 score | –0.032*** | –0.001 | 0.033*** | 0.016 | 0.004 | 0.043*** | 0.045*** | 0.004 | 0.034*** |
| | (0.012) | (0.013) | (0.012) | (0.013) | (0.011) | (0.012) | (0.013) | (0.012) | (0.013) |
| Risk measure | –0.031 | –0.008 | 0.022 | 0.005 | 0.024 | –0.001 | –0.001 | 0.021 | 0.019 |
| | (0.025) | (0.022) | (0.023) | (0.023) | (0.021) | (0.022) | (0.024) | (0.023) | (0.023) |
| Competitiveness | 0.223** | 0.291*** | 0.189** | 0.237*** | 0.187* | 0.090 | 0.076 | 0.309*** | 0.247*** |
| | (0.092) | (0.090) | (0.093) | (0.091) | (0.096) | (0.088) | (0.096) | (0.088) | (0.090) |
| Self-esteem | 0.019 | –0.014 | –0.022 | 0.012 | –0.027** | –0.025* | –0.017 | –0.005 | –0.015 |
| | (0.014) | (0.013) | (0.014) | (0.014) | (0.013) | (0.014) | (0.015) | (0.014) | (0.014) |
| Mother works | –0.286** | –0.251** | –0.299** | –0.333*** | –0.168 | 0.040 | –0.013 | –0.026 | 0.005 |
| | (0.138) | (0.123) | (0.127) | (0.125) | (0.128) | (0.128) | (0.131) | (0.130) | (0.131) |
| Mother has UG degree | 0.230 | 0.021 | –0.142 | –0.163 | –0.229 | 0.140 | 0.094 | 0.183 | 0.247 |
| | (0.186) | (0.183) | (0.156) | (0.173) | (0.163) | (0.187) | (0.189) | (0.189) | (0.180) |
| Year effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 322 | 322 | 322 | 322 | 322 | 322 | 322 | 322 | 322 |
| Log pseudolikelihood | -347.36 | -419.23 | -422.30 | -401.40 | -428.23 | -431.95 | -378.51 | -421.76 | -406.99 |

# 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?

# Sensitivity of choice to parameters

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

| | Science | | | Business | | | Humanities | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) All | (2) Men | (3) Women | (4) All | (5) Men | (6) Women | (7) All | (8) Men | (9) Women |
| Love of subject | -0.12*** | -0.09 | -0.11* | -0.07* | -0.06 | -0.09* | 0.19*** | 0.15*** | 0.20*** |
| | (0.03) | (0.05) | (0.05) | (0.03) | (0.04) | (0.04) | (0.03) | (0.04) | (0.05) |
| Grades | -0.07* | -0.05 | -0.06 | 0.03 | 0.02 | 0.03 | 0.04 | 0.03 | 0.04 |
| | (0.03) | (0.04) | (0.04) | (0.03) | (0.04) | (0.04) | (0.02) | (0.03) | (0.04) |
| Wages | 0.04 | -0.04 | 0.10* | 0.05 | 0.05 | 0.07 | -0.10*** | -0.01 | -0.17*** |
| | (0.03) | (0.04) | (0.04) | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.04) |
| Career | -0.03 | -0.01 | -0.04 | 0.04 | 0.05 | 0.01 | -0.01 | -0.04 | 0.03 |
| | (0.03) | (0.04) | (0.05) | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.04) |
| Work-life balance | 0.03 | 0.06 | 0.03 | -0.03 | -0.04 | -0.03 | -0.01 | -0.02 | 0.01 |
| | (0.03) | (0.04) | (0.04) | (0.03) | (0.03) | (0.04) | (0.02) | (0.03) | (0.03) |
| Parents | 0.04 | 0.04 | 0.03 | -0.03 | -0.00 | -0.05 | -0.01 | -0.04 | 0.02 |
| | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.03) | (0.02) | (0.02) | (0.04) |
| Teacher | -0.03 | -0.07 | -0.02 | -0.03 | 0.01 | -0.05 | 0.06*** | 0.06** | 0.08** |
| | (0.03) | (0.04) | (0.04) | (0.03) | (0.04) | (0.03) | (0.02) | (0.02) | (0.03) |
| Peers | 0.10** | 0.17*** | 0.05 | -0.02 | -0.08 | 0.02 | -0.07** | -0.09** | -0.07 |
| | (0.03) | (0.04) | (0.04) | (0.03) | (0.04) | (0.03) | (0.03) | (0.03) | (0.04) |
| Option value | -0.01 | 0.01 | -0.03 | 0.05* | 0.01 | 0.10** | -0.04* | -0.01 | -0.07* |
| | (0.02) | (0.03) | (0.04) | (0.02) | (0.03) | (0.04) | (0.02) | (0.02) | (0.03) |
| Observations | 326 | 150 | 176 | 326 | 150 | 176 | 326 | 150 | 176 |

# 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

# Results:summary

**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