# The gender pay gap amongst high educated workers: evidence from academia

Rachel Griffith
with
Karim Chalak (Manchester)
and
John Gathergood (Nottingham)

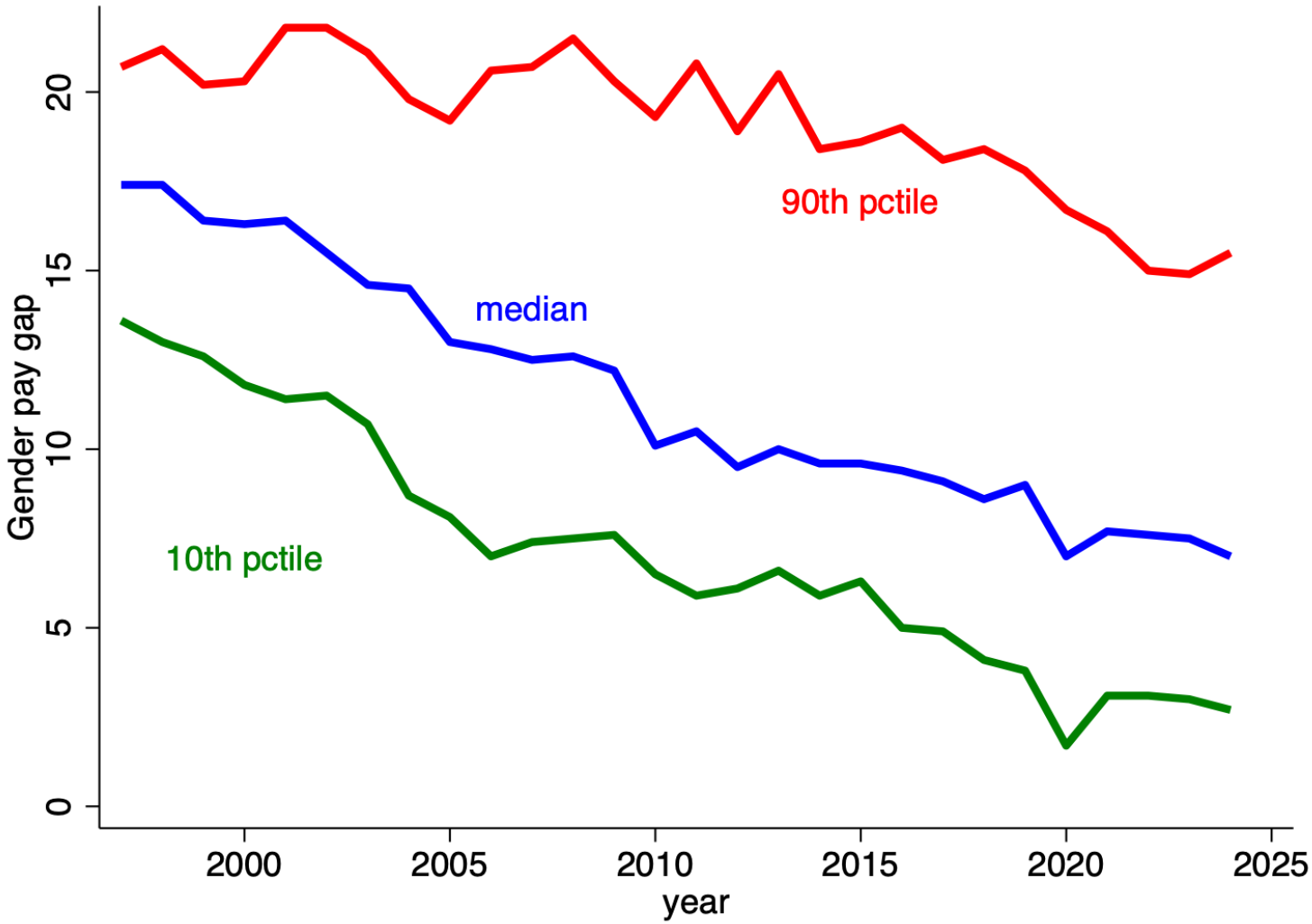
April 2025

Institute for Fiscal Studies

MANCHESTER 1824

# The gender pay gap is highest for high earners

## Whole UK Economy




Source: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/genderpaygapintheuk/2024

# Motivation

- Reductions in the gender pay gap over the past 25 years due to
  - increased education attainment by women, who are now on average more highly educated than men
  - increased minimum wages, which has brought down the gender pay gap at low wages
- But there has not been comparable progress for highly educated women, where gender pay gaps remain large
- Gender gaps in pay have substantial consequences for inequalities
- And suggest that the talents of women are not being used in the most productive way possible

# Motivation

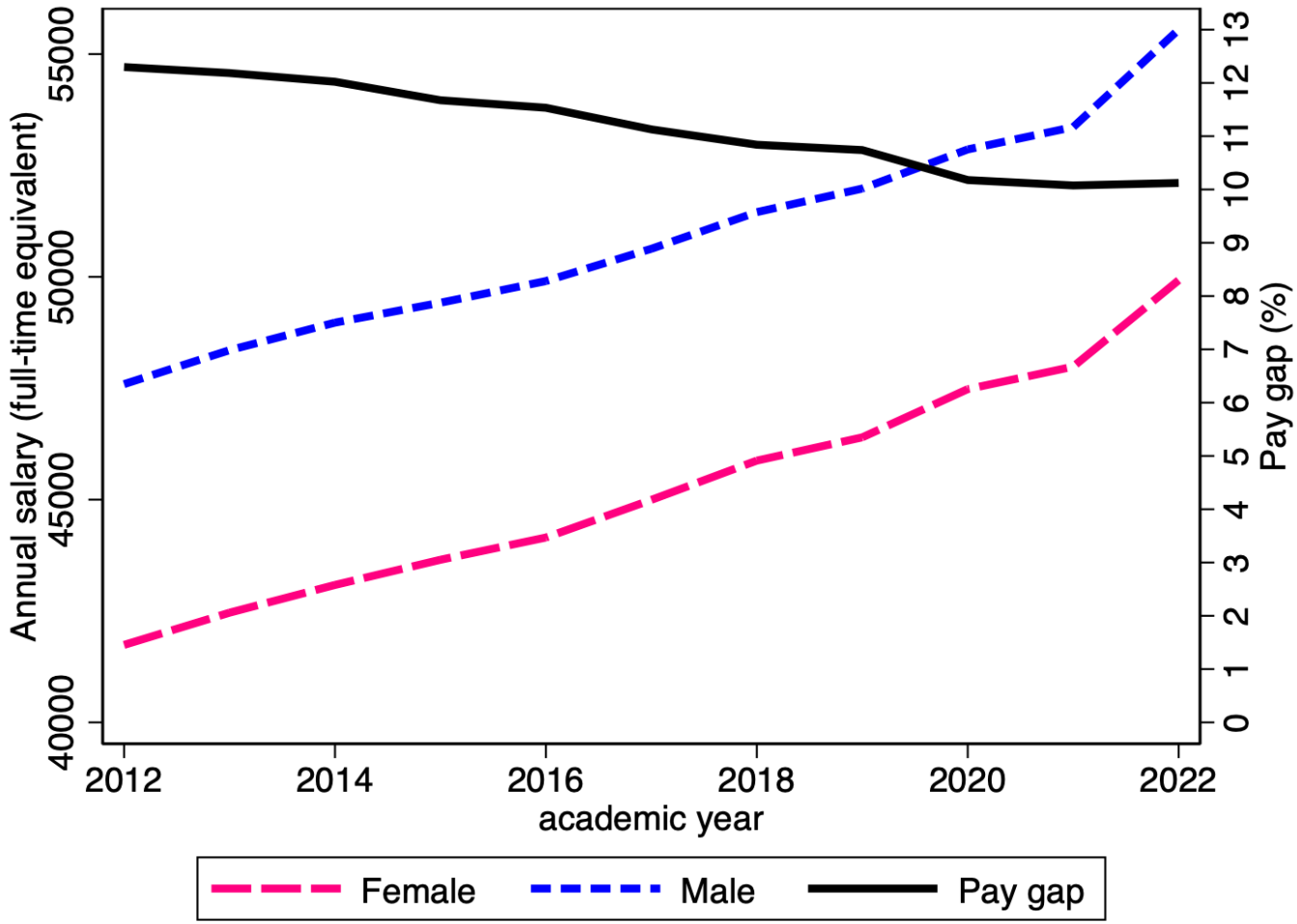
- The literature points to the large gender gaps arising in occupations where there are higher returns for longer/less flexible hours (Goldin, Bertrand and others)
  - if women place higher value on time at home with young children then they may be willing to trade lower pay for more flexible work hours
- Some occupations have a non-linear or convex wage structure, where workers are not close substitutes for each other, so there is high demand (and compensation) for an individual's time
  - e.g. trial lawyers and consultants are occupations that require long hours, where workers are not close substitutes for each other, pharmacists are close substitutes so hours can be more flexible
- Women do less well in occupations that require long and inflexible hours to remain on the "fast track", because they are difficult to combine with family commitments

# Our contribution

- We study the determinants of the gender pay gap amongst academics in the UK
- This is an interesting setting because
  - many high paid workers who have all invested heavily in human capital
  - large and persistent gender pay gaps in some disciplines but not in others
  - many people doing a similar job, but requirements for success and the reward structures differ across disciplines
  - we have high quality data on pay and outputs for the population of workers
  - we can identify research active academics
  - an interesting industry in itself

# Large and persistent gender pay gap

all academic (teaching and research) staff, all disciplines, all institutions

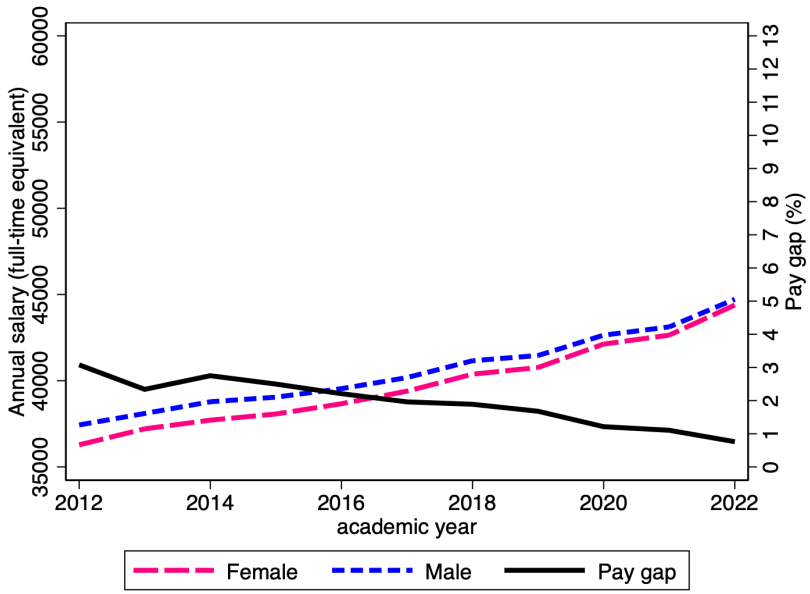


# Data on pay

- Salary of all academic staff in UK higher education institutions
  - from Higher Education Statistics Agency (HESA) staff records
  - academic years 2012-13 - 2022-23
  - all institutions, all disciplines
  - around 200,000 individuals in each year
  - salary, hours and contract (FTE, full/part-time; permanent/fixed term)
  - function: teaching, research, teaching and research (T&R)
  - discipline, institution, age, gender, ethnicity
  - parental leave
  - whether submitted to REF2014 and REF2021

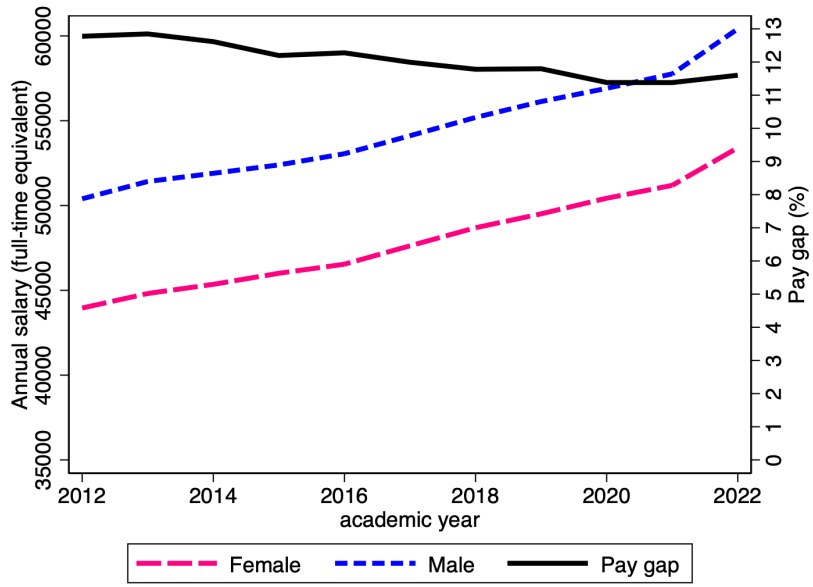
# The pay gap arises in research

Teaching




32% of staff on teaching contracts

Research (inc T&R)

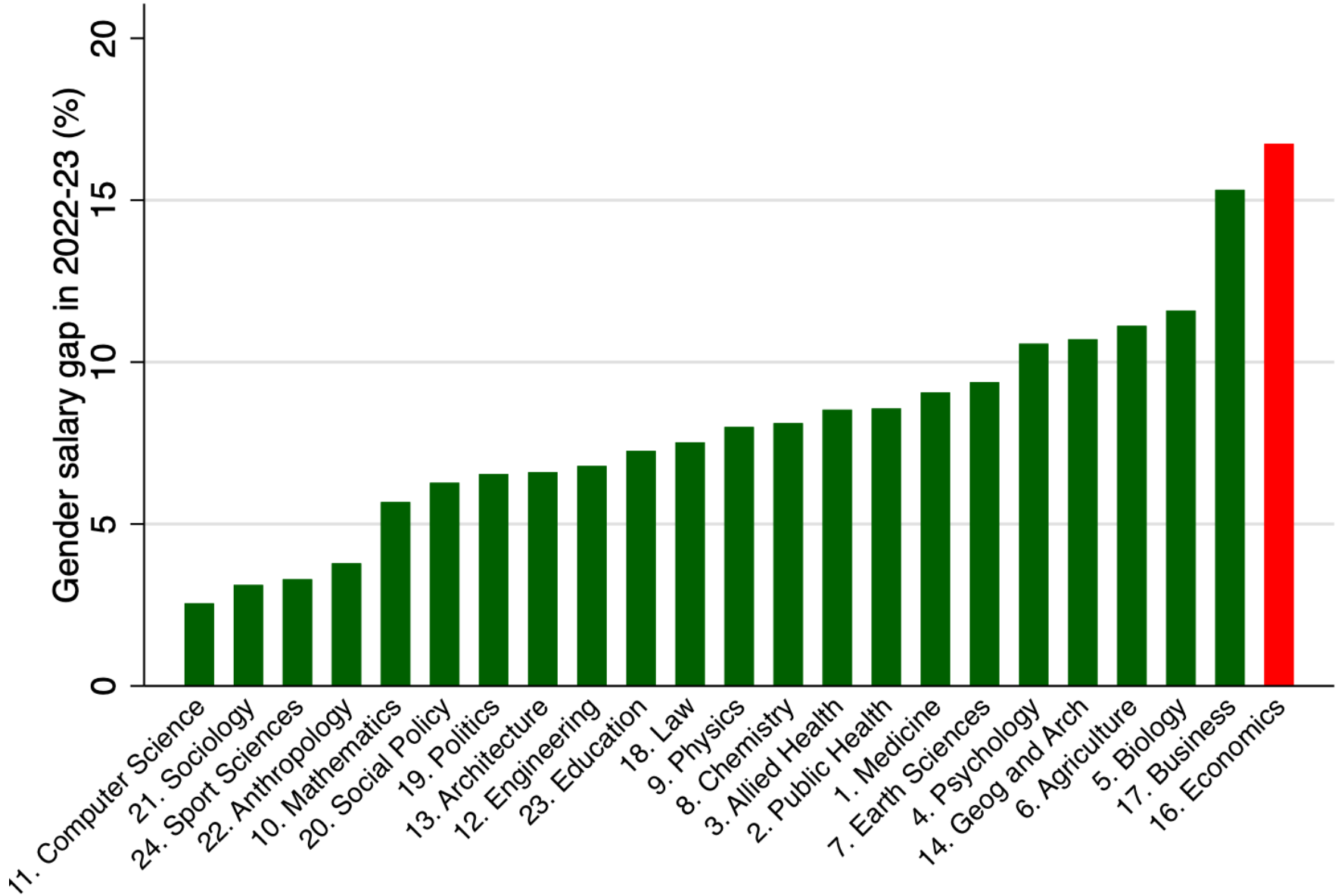



68% on research contracts

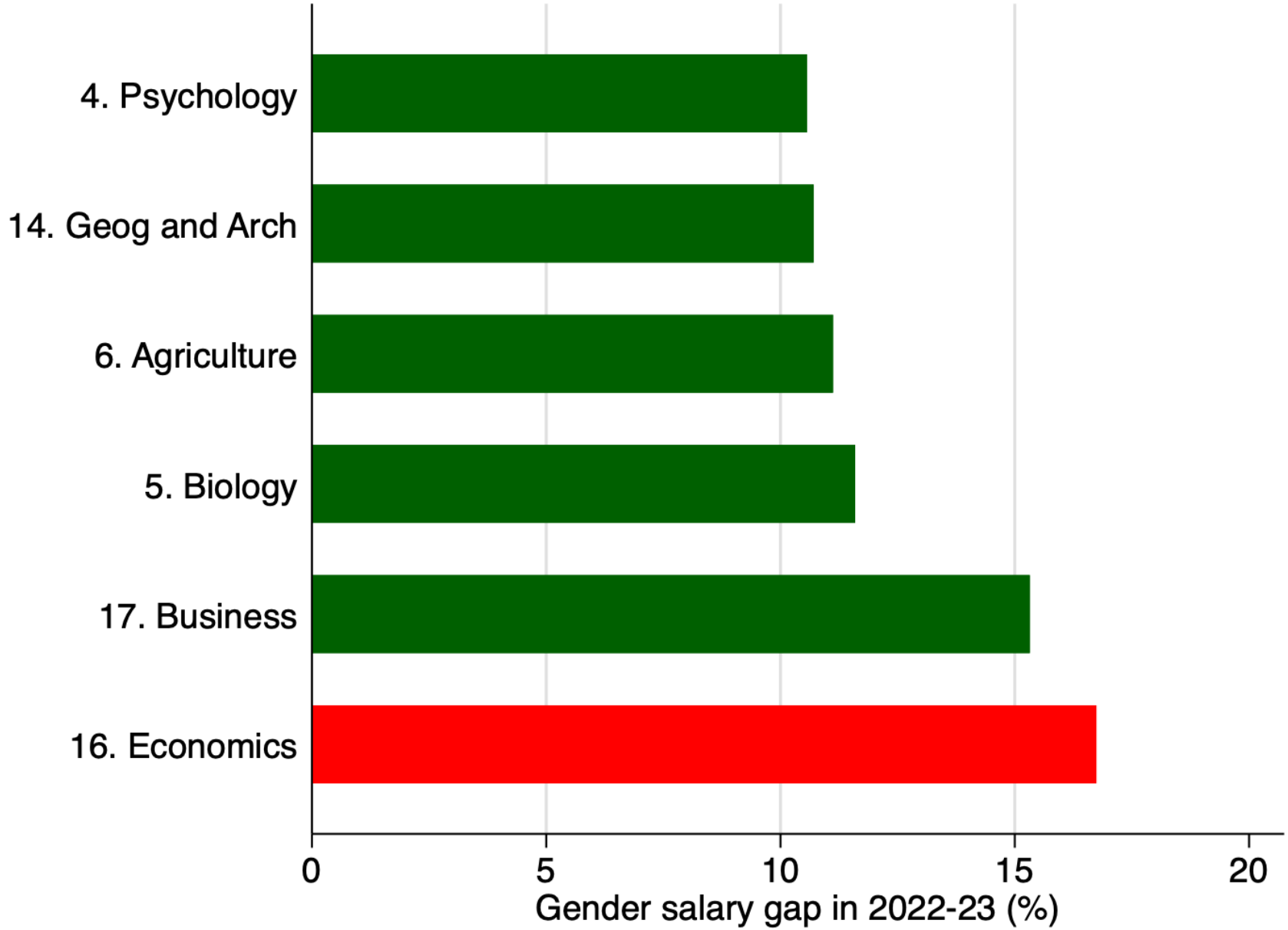
# We study research active academics

- academics submitted to REF2021 or REF2014
- work in one of 56 research oriented institutions (research income is $>$15% income)
- work in one of 23 disciplines that publish in journals
  - includes all of Panels A (Medicine, Health, Life Sciences), B (Physical Sciences, Engineering, Maths), and C (Social Sciences)
  - excludes Panel D (Arts and Humanities)
- around 35,000 individuals submitted to REF2014
- around 45,000 individuals submitted to REF2021
- around one-third are female

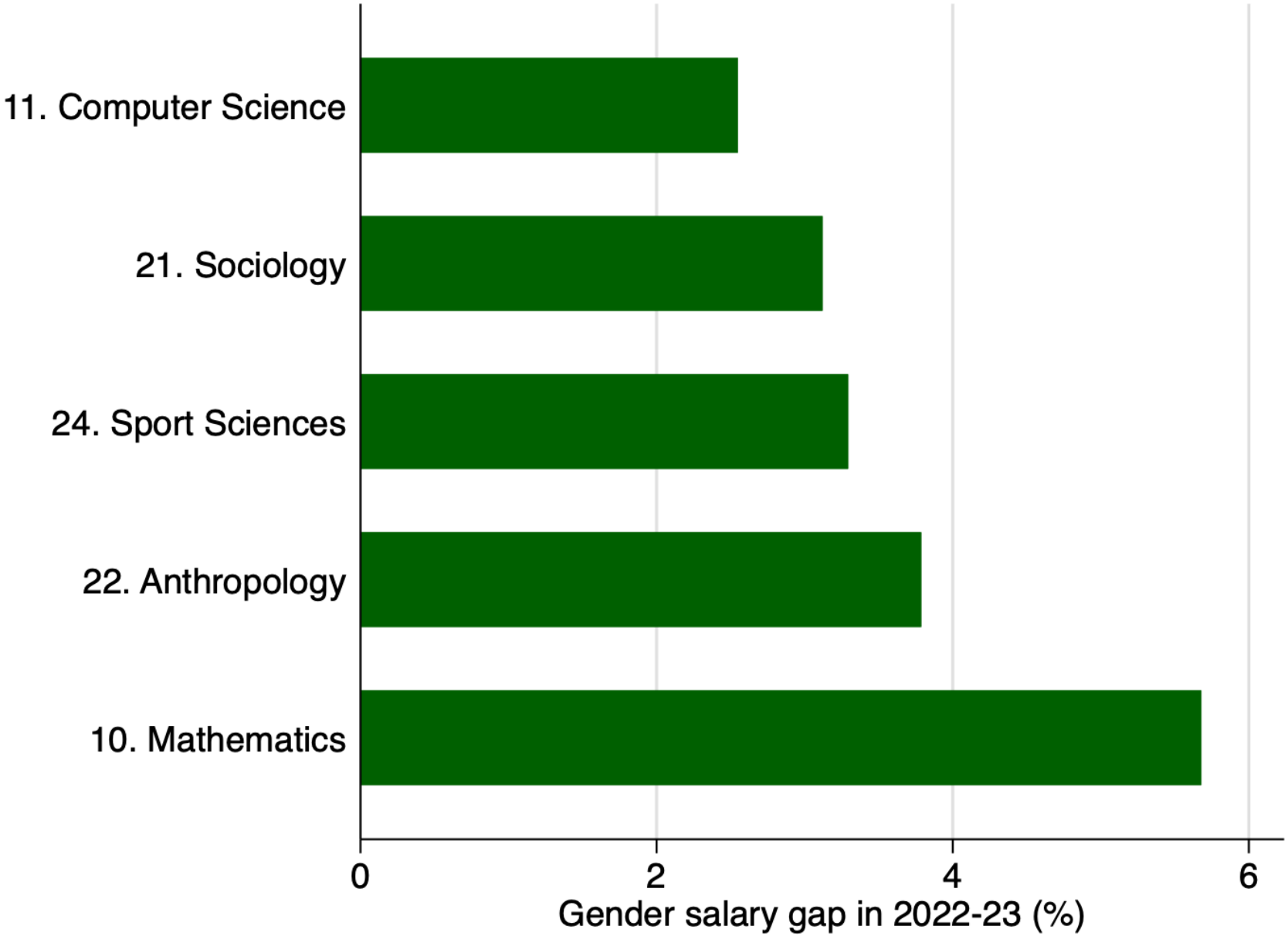
# Gender pay gap varies across discipline



# Highest gender pay gap



# Lowest gender pay gap



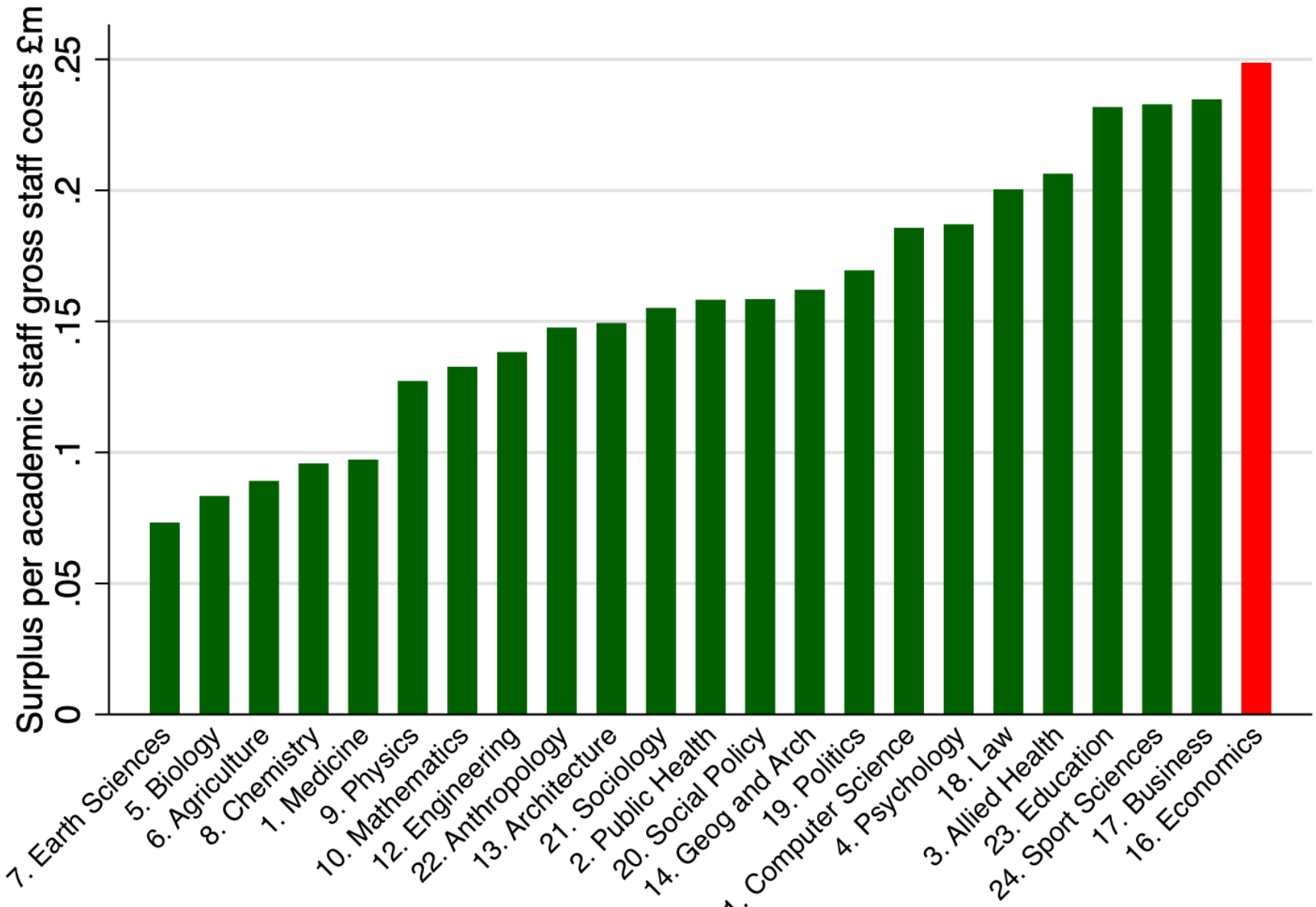
# The wage setting process in academia

- Universities earn revenue from high quality research
  - high quality research attracts research funding and overseas students
  - the surplus that a university can generate from high quality research differs by discipline depending on costs, attractiveness of course, etc.
- Researchers differ in their productivity in producing high quality research
  - individuals differ in their ability and effort to produce high value outputs
  - disciplines differ in the research expectations and reward structures, with some having large returns to “big” publications, others having more incremental structure of reward, ...
- Researchers and the university bargain over pay, depending on
  - researchers’ outside option
  - individual differences in mobility, risk preferences, ...

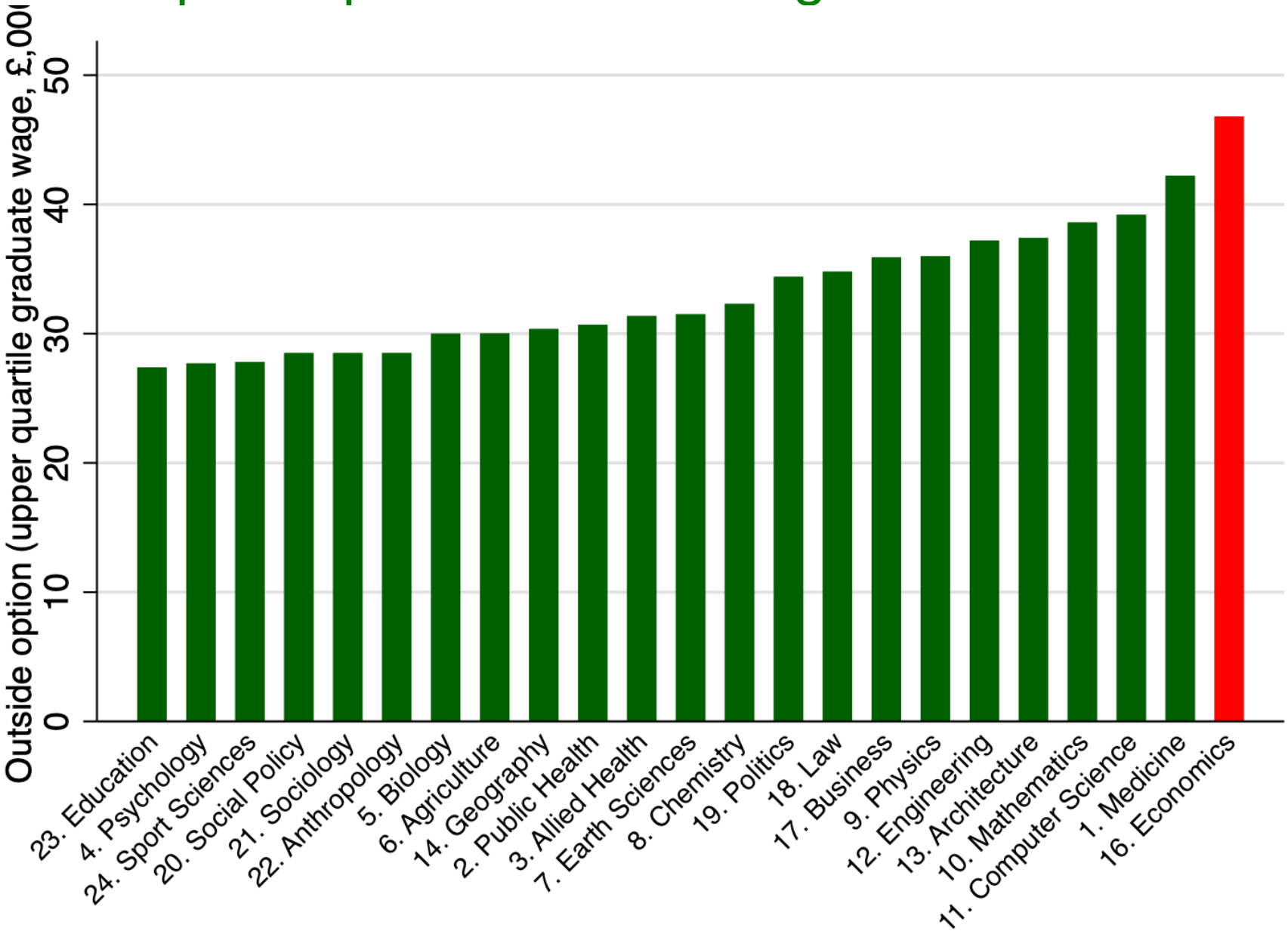
# Data on disciplines

- Surplus
  - from HESA finance, student and staffing records
  - total revenue (tuition + research) - total variable costs (excl academic salary) per academic staff
- Outside option
  - Longitudinal Educational Outcomes (LEO): matches HMRC (tax records) earnings and employment data with HESA student records
    - earnings for students in each subject area five years after graduation (UK domiciled first degree graduates from HEIs in Great Britain, 2019/20 tax year)
  - Share of staff that are international
- Publication norms and expectations
  - Effort required to get a high valued publication
  - concentration of publications by top academics (staff in top 3 departments by REF2021 GPA), compared to publications by staff outside top 3
  - average page length, time to publish, rejection rates, number of publications per person per year

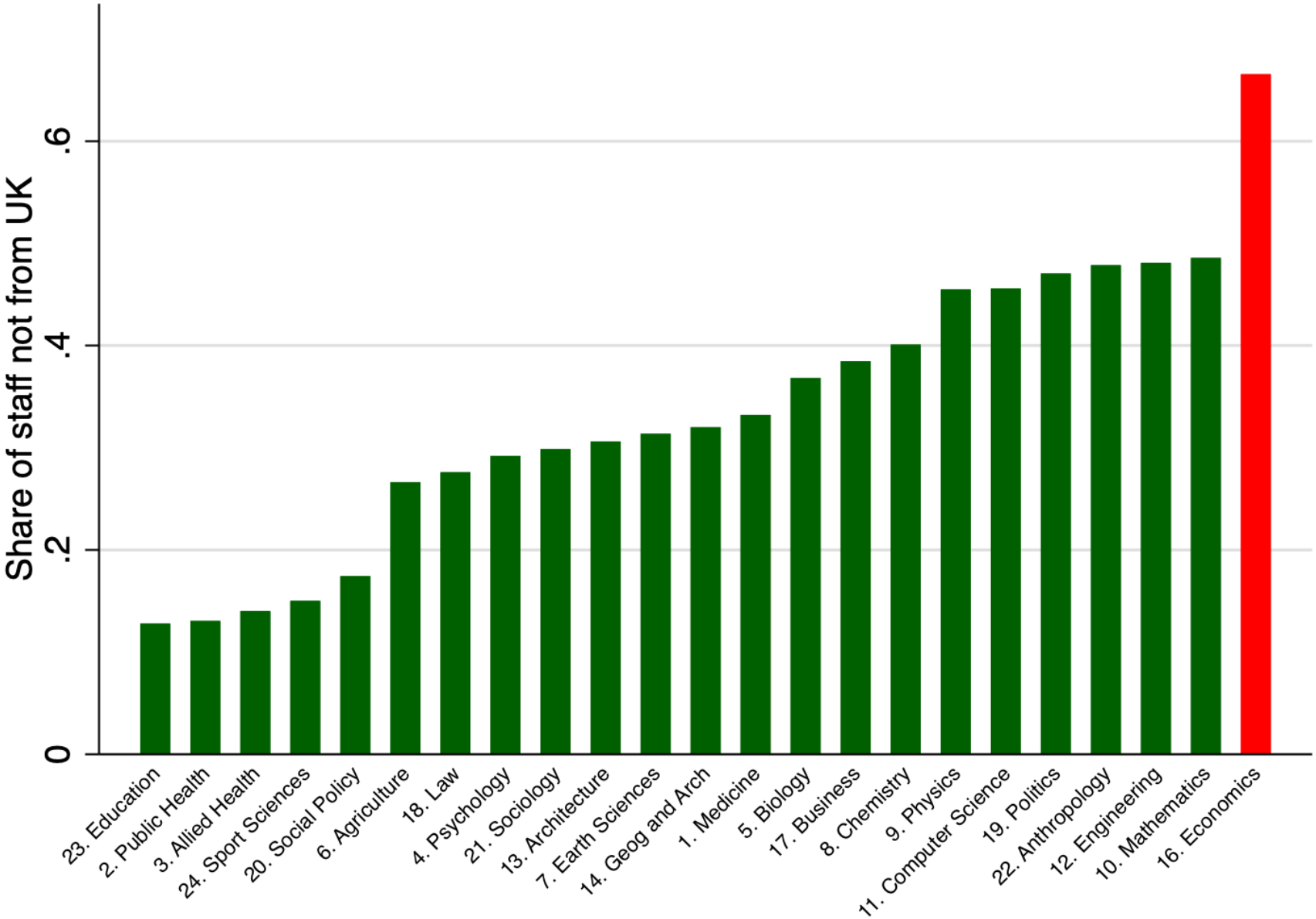
# Surplus per academic staff



# Outside option: private sector wage



# Outside option: % staff that are international

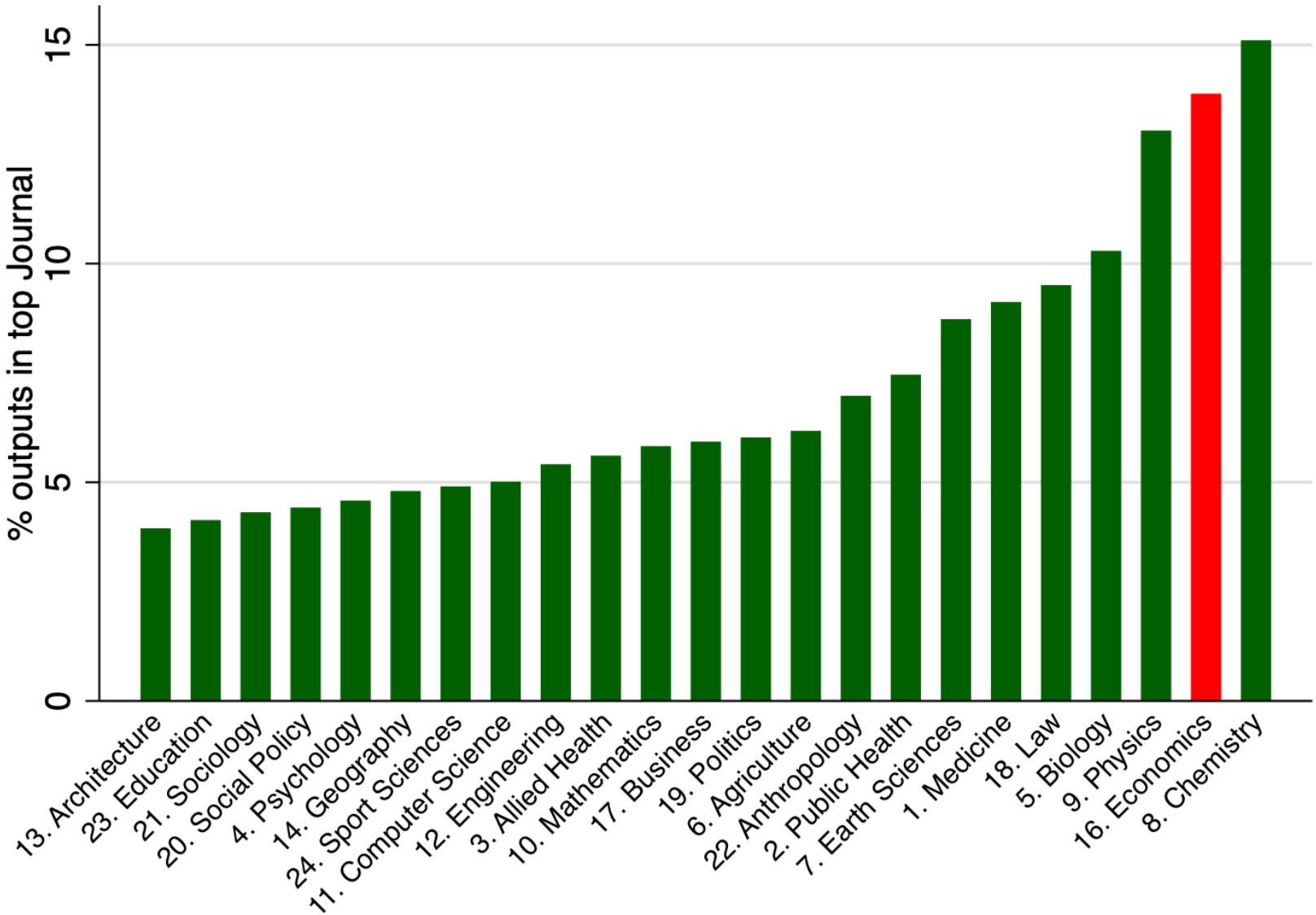


# Publication norms: the top Journal

- Consider all papers submitted to REF2021
  - what % were submitted to each Journal
  - compare % in departments that got the highest % of 4* grades and All departments
- In Economics the Journal that is most common amongst the top 3 departments is **The American Economic Review**

# % REF2021 outputs in top journal

in top 3 departments

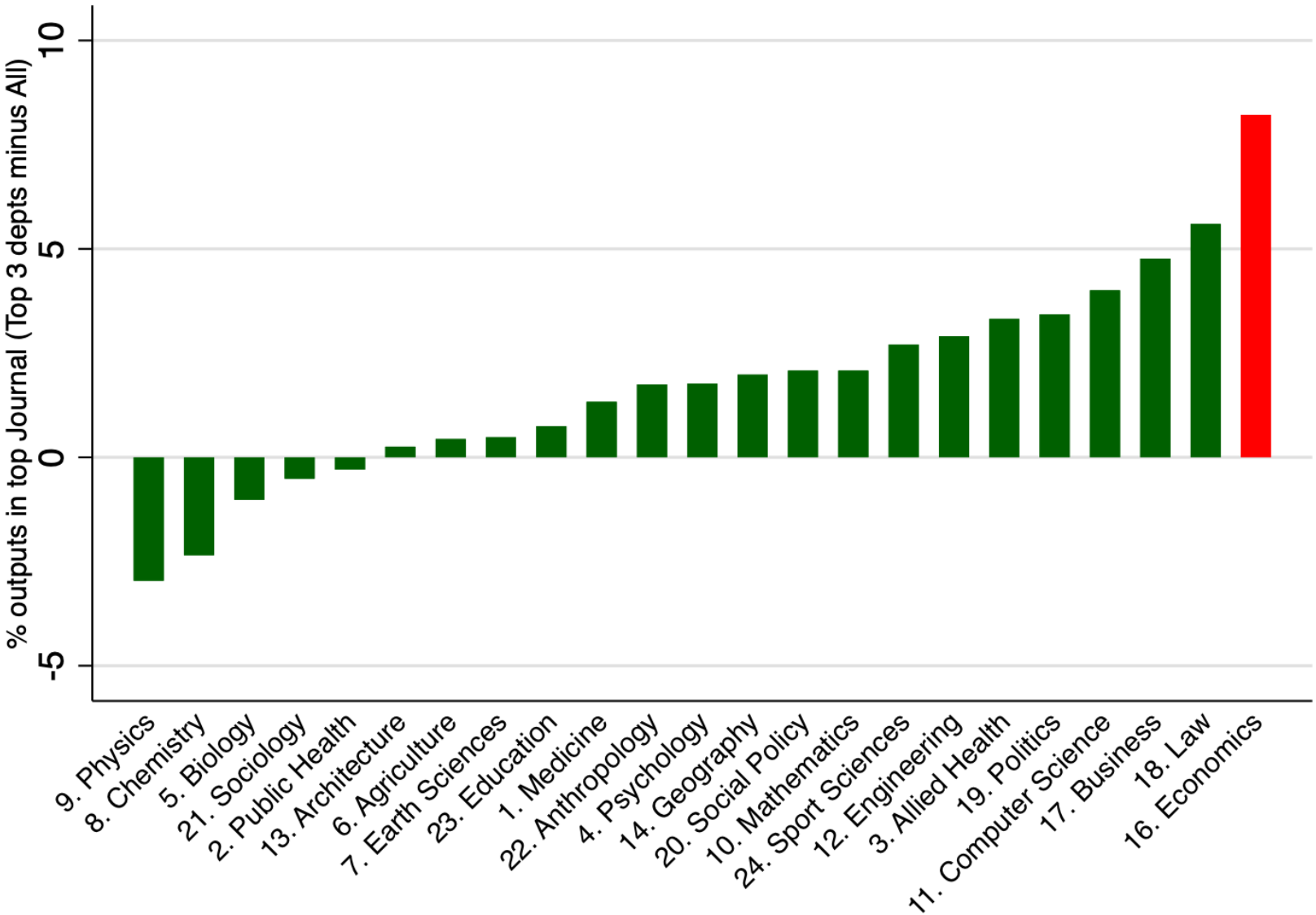


# Publication norms - the top Journal

- In contract, in economics if we look at submissions from all departments to **The American Economic Review**
- they represent 5.7% of submissions
- compared to 13.9% in the top 3 deptments
- a difference of 8.2 p.p.

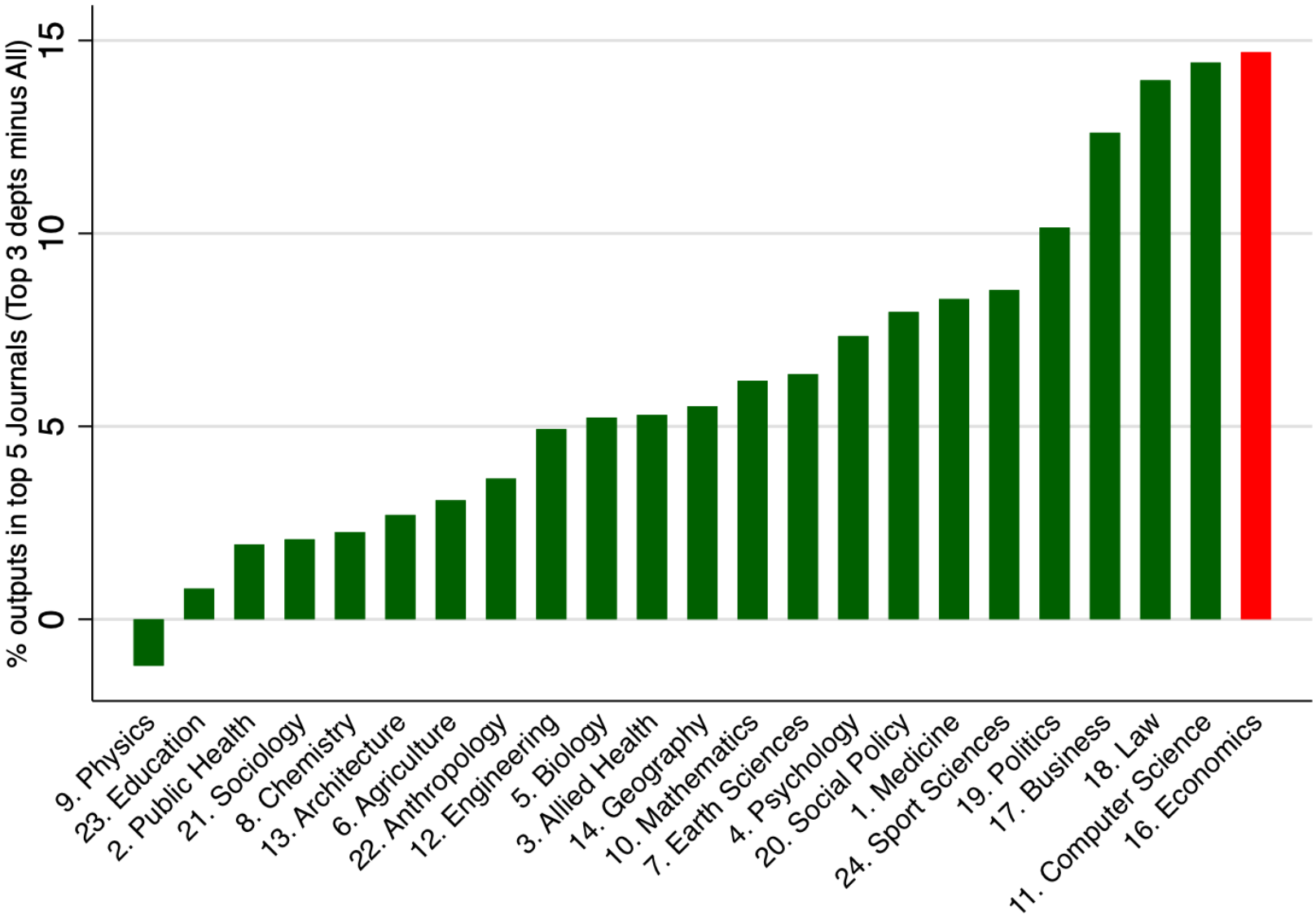
# Difference in % outputs in leading journal

## Top 3 departments minus All



# Difference in % outputs in Top 5 Journals

## Top 3 departments minus All



# Publication norms

In addition, large differences across journals in:

- rejection rates, decision times, article page lengths, number of articles published per year

| UoA | Highest % journal | Top 3 | All | diff |
|---|---|---|---|---|
| 8. Chemistry | Journal Of The American Chemical Society | 15.1 | 17.5 | -2.3 |
| 16. Economics | American Economic Review | 13.9 | 5.7 | 8.2 |
| 9. Physics | Physical Review Letters | 13.0 | 16.0 | -3.0 |
| 5. Biology | Nature Communications | 10.3 | 11.3 | -1.0 |
| 18. Law | Oxford Journal Of Legal Studies | 9.5 | 3.9 | 5.6 |
| 1. Medicine | Nature Communications | 9.1 | 7.8 | 1.3 |

# Data on individual research outputs

- We have the universe of research active academics
- We construct the universe of their publications and citations
  - take list of outputs submitted to REF2021 and REF2014
    - 89% of outputs have a DOI, 91% of these in Scopus
  - use Scopus to identify the submitting author(s)
    - get the institutional affiliation for all authors and match to the institution that submitted the output
    - get the full publication histories of all these authors from Scopus
    - use Gender Guesser and GenderIO to identify whether male or female
    - year of first publication gives an estimate of age
- Construct total citations, H-index, number in top journal, top 5, top 20 journals, and other measures of outputs

# How important are outputs for salary

We are interested in learning about features of the joint distribution of output, salary and other characteristics

- for example, $\beta$, the elasticity of salary wrt output

$$Y_i = X_i'\beta + W_i'\gamma + e_i$$

$i$: individuals
$Y_i$: salary
$X_i$: vector of outputs
$W_i$: vectors of individual characteristics observed in both salary and output data
$e_i$: idiosyncratic error

- We observe $(Y_i, W_i)$ for the population
- We observe $(X_i, W_i)$ for the population
  - $W$: discipline, institution, gender, age

# Imputing outputs into salary data

- We impute a value of outputs, $\tilde{X}_i$, for every individual in the salary data
  - average of output over nearest (age-wise) neighbours who share same (female, institution, discipline)
- We observe actual age in salary data
- In output data we observe year of first publication
  - we estimate age assuming a minimum age at year of first publication that varies by discipline (but not within discipline)
  - we select this minimum age to match a set of moments in the salary and outputs data (means, variances and covariances of $X$, $W$, age)

# Empirical model of pay

$$log(Y_i) = \tilde{X}_i \beta_d + W_i' \phi_d + Z_i' \gamma_d + e_i$$

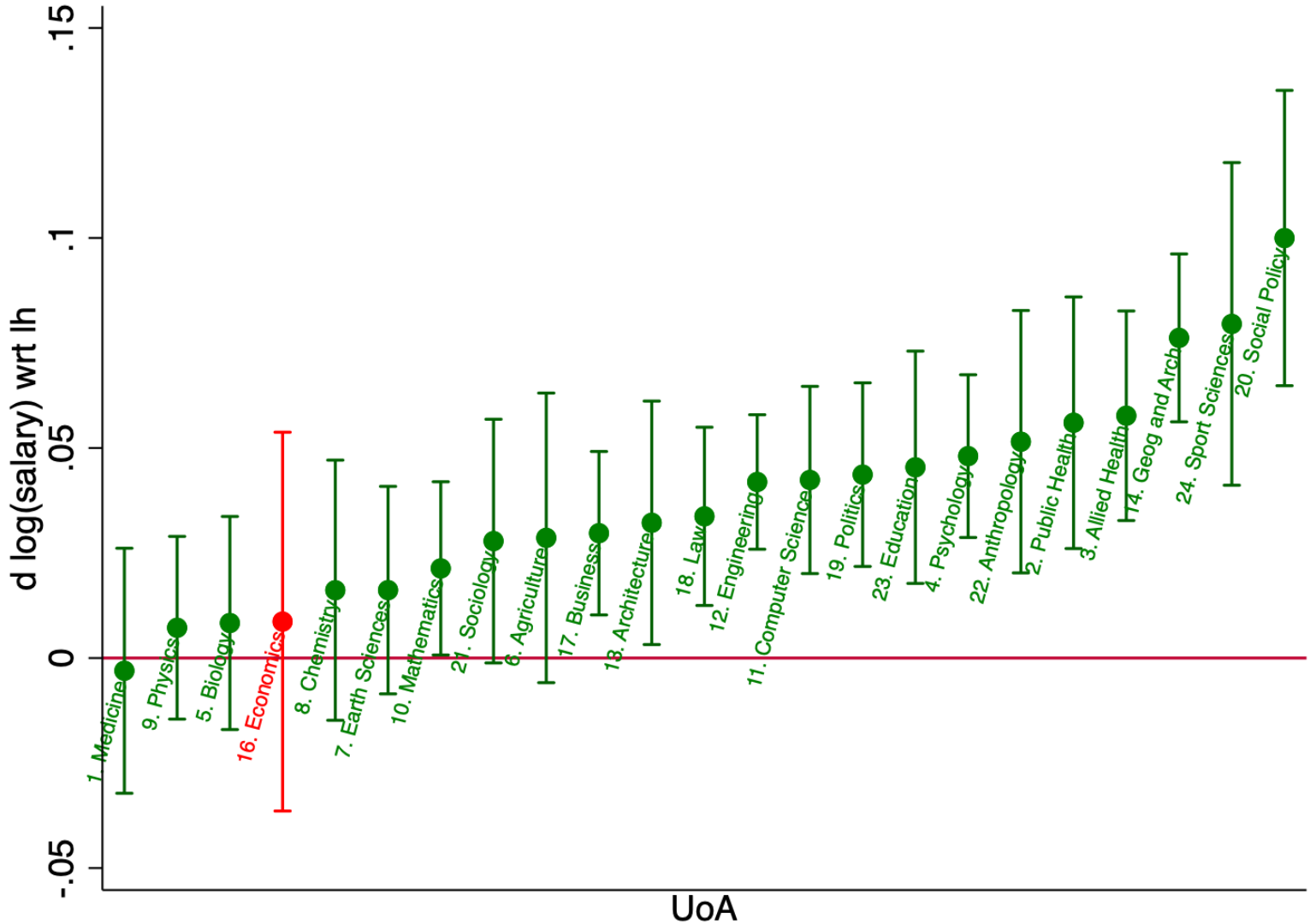
- run separately for each discipline (d)
- W: common covariates (age, gender), Z: covariates just in Y data
- We are interested in
  - estimates of $\beta$, the elasticity of salary wrt output
  - how important differences in outputs ($X$) are to explain the gender pay gap:

$$\left(\overline{Y^F} - \overline{Y^M}\right)_d = \left(\overline{X^F} - \overline{X^M}\right)_d \widehat{\beta}_d + \left(\overline{W^F} - \overline{W^M}\right)_d \widehat{\phi}_d + \left(\overline{Z^F} - \overline{Z^M}\right)_d \widehat{\gamma}_d$$

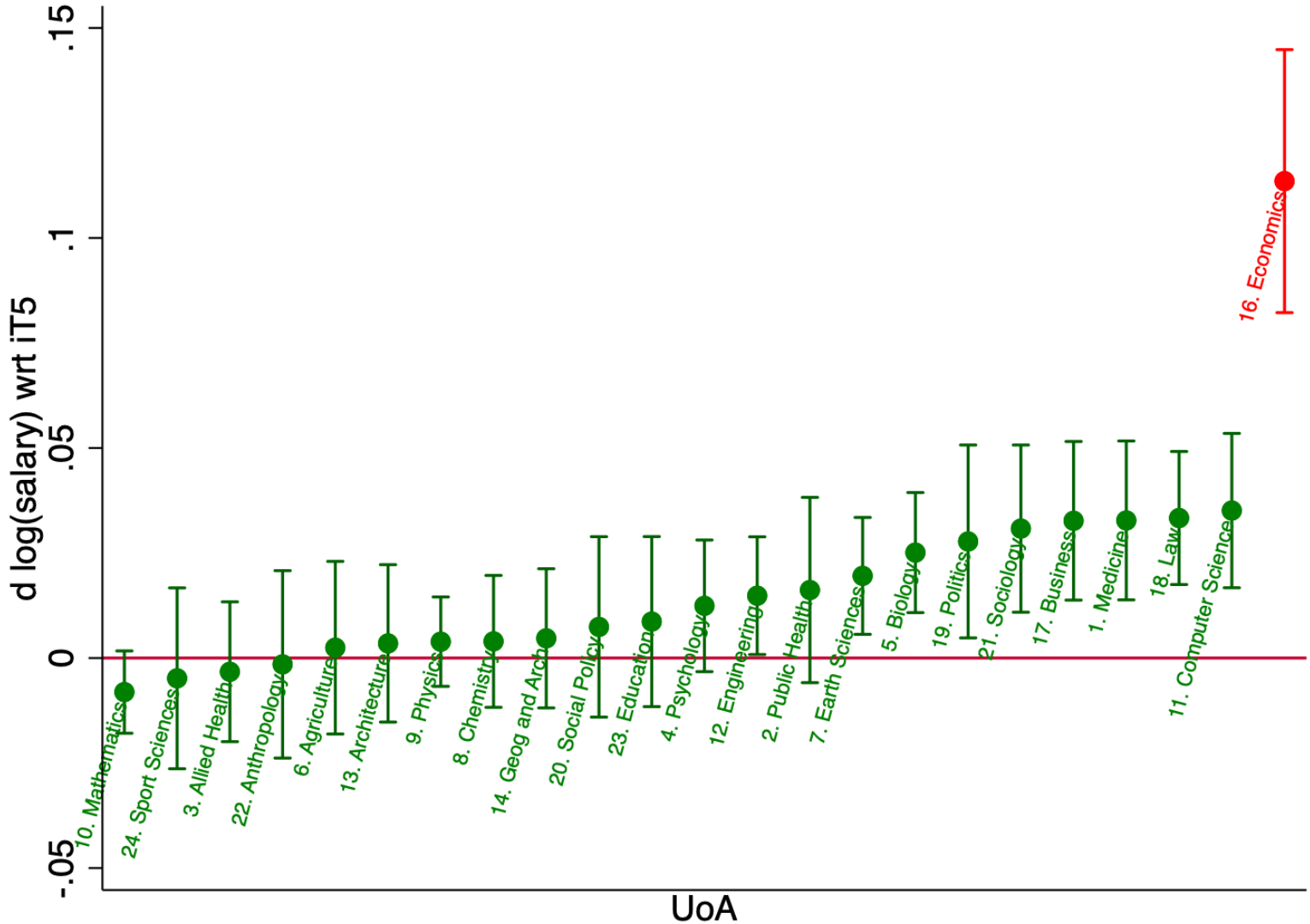
where $\overline{X^F}$ and $\overline{X^M}$ denote the mean of a variable across females and males

$$\frac{\left(\overline{X^F} - \overline{X^M}\right) \widehat{\beta}}{\left(\overline{Y^F} - \overline{Y^M}\right)}$$

# Elasticity of salary wrt H-Index

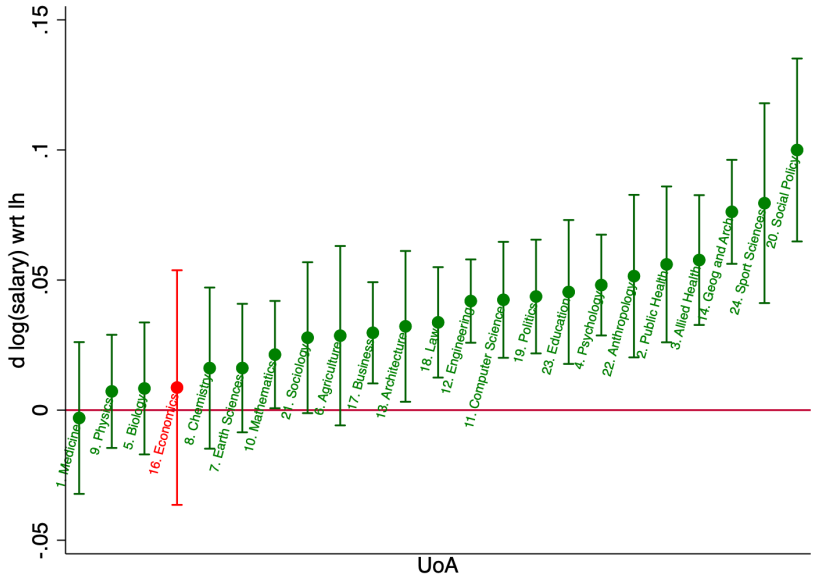


# Elasticity of salary wrt Top 5

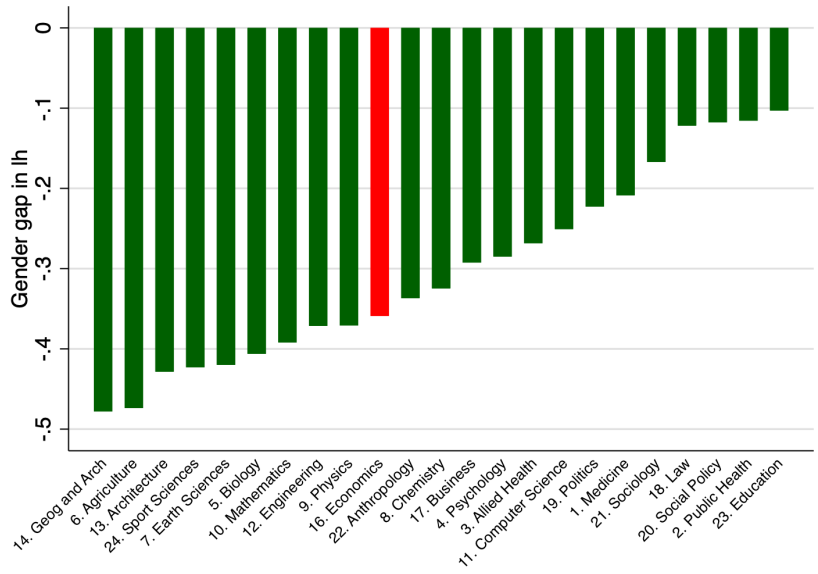


# Contribution of H-index

Elasticity of salary wrt H-index



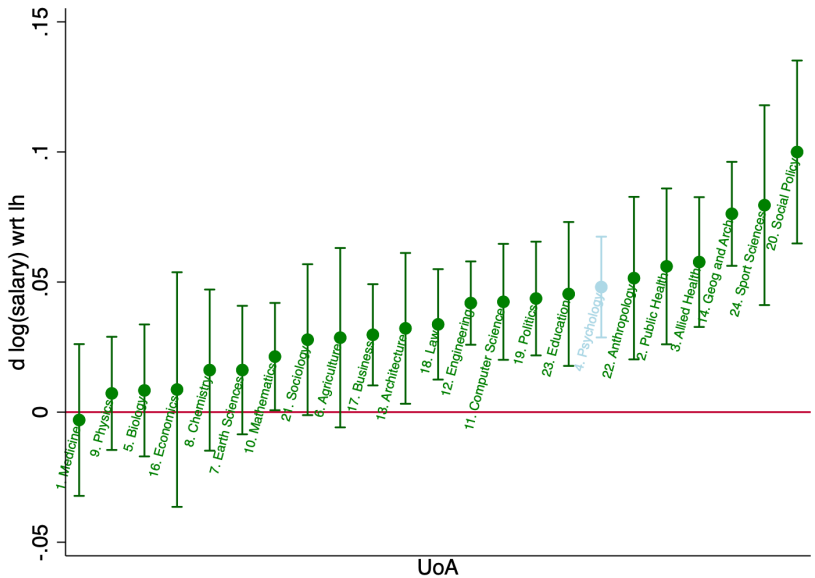

Gender difference in H-index



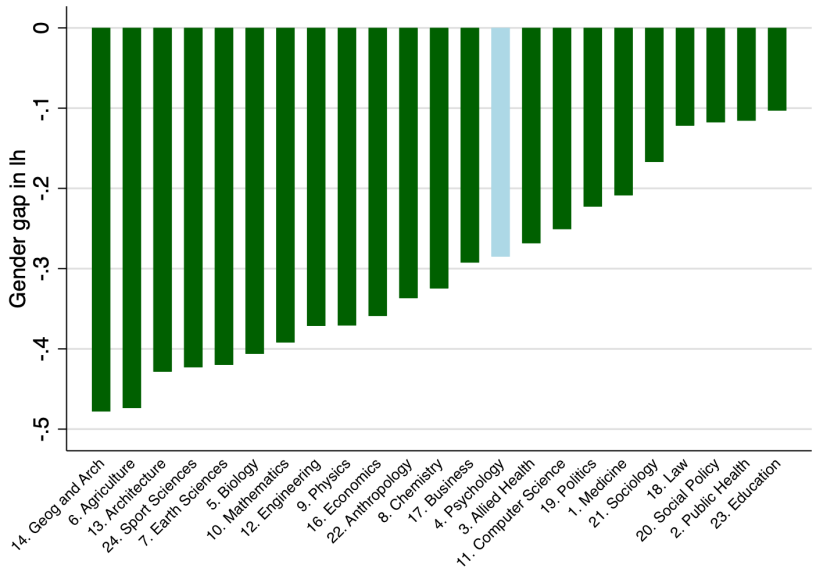

- the elasticity of salary wrt to H-index is low in economics
  - so gender differences in H-index don't contribute much to gender pay gap

# Contribution of H-index

Elasticity of salary wrt H-index



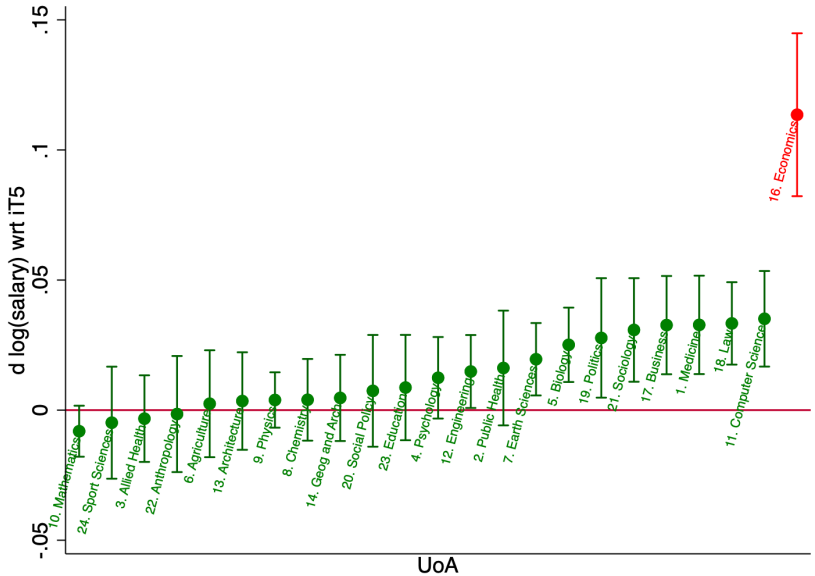

Gender difference in H-index



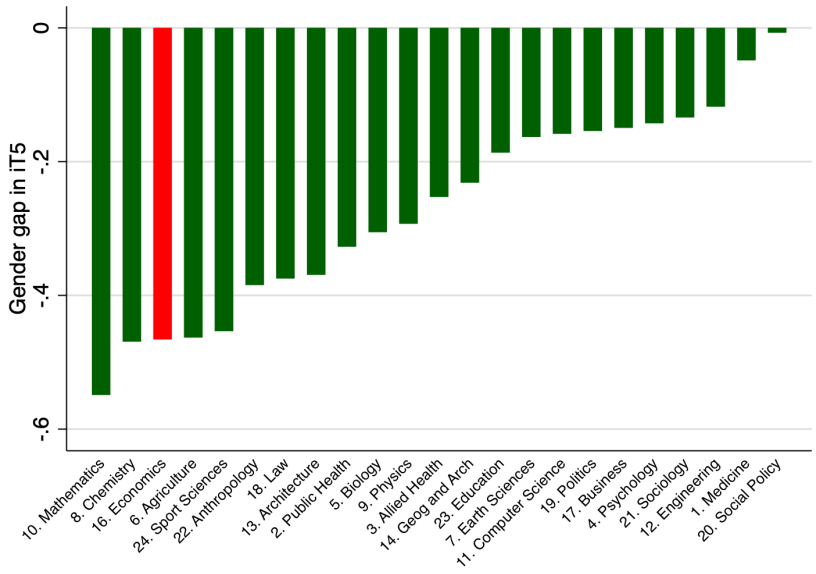

- the elasticity of salary wrt to H-index is higher in psychology
  - so lower gender differences in H-index contribute more to gender pay gap

# Contribution of Top 5

Elasticity of salary wrt Top 5



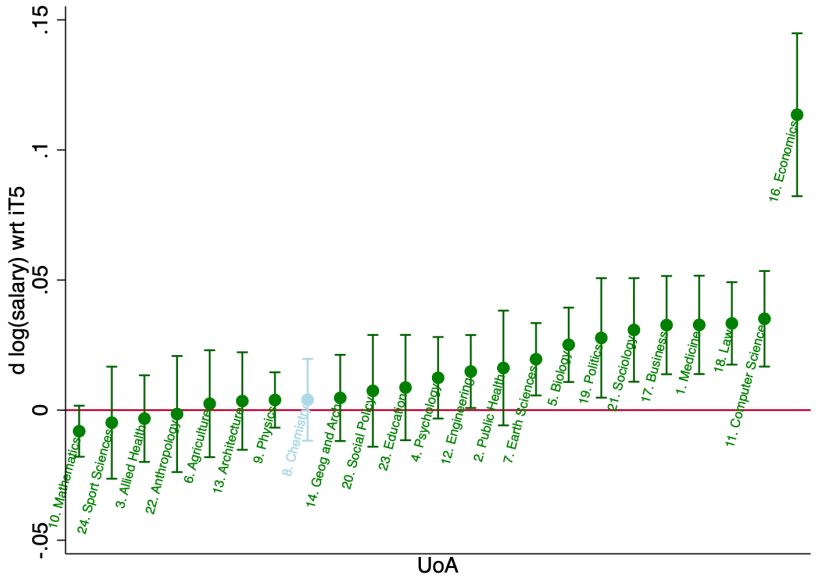

Difference in Top 5



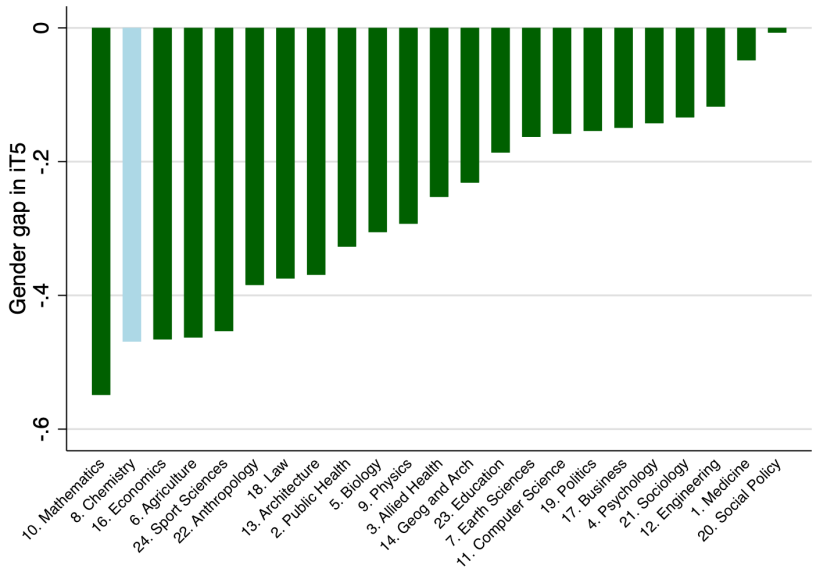

- the elasticity of salary wrt to Top 5 publications is high in economics
  - so gender differences in Top 5 contribute to the gender pay gap

# Contribution of Top 5

Elasticity of salary wrt Top 5




Difference in Top 5

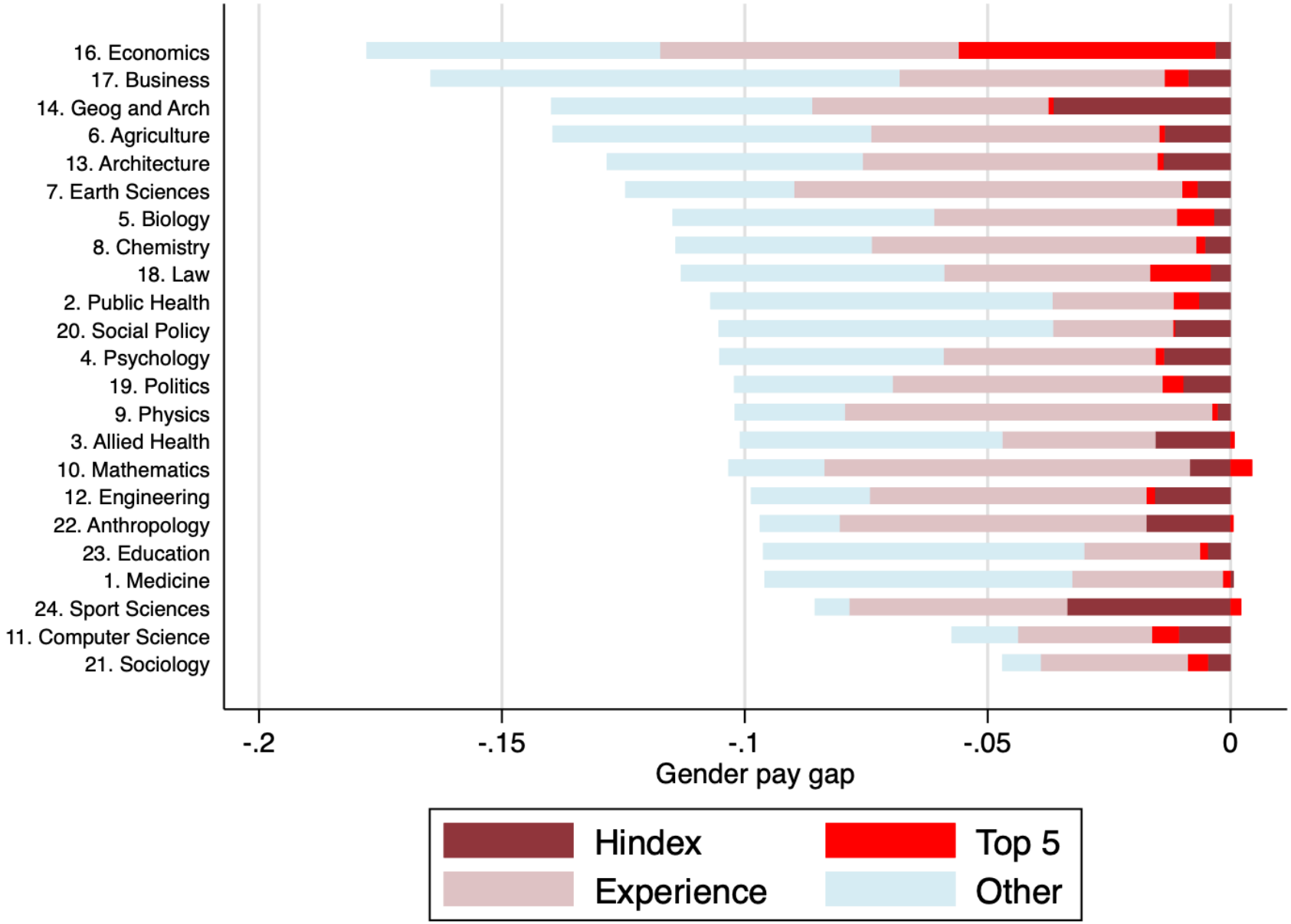



- the elasticity of salary wrt to Top 5 is lower in chemistry
  - so a higher gender differences in Top 5 does not contribute to gender pay gap

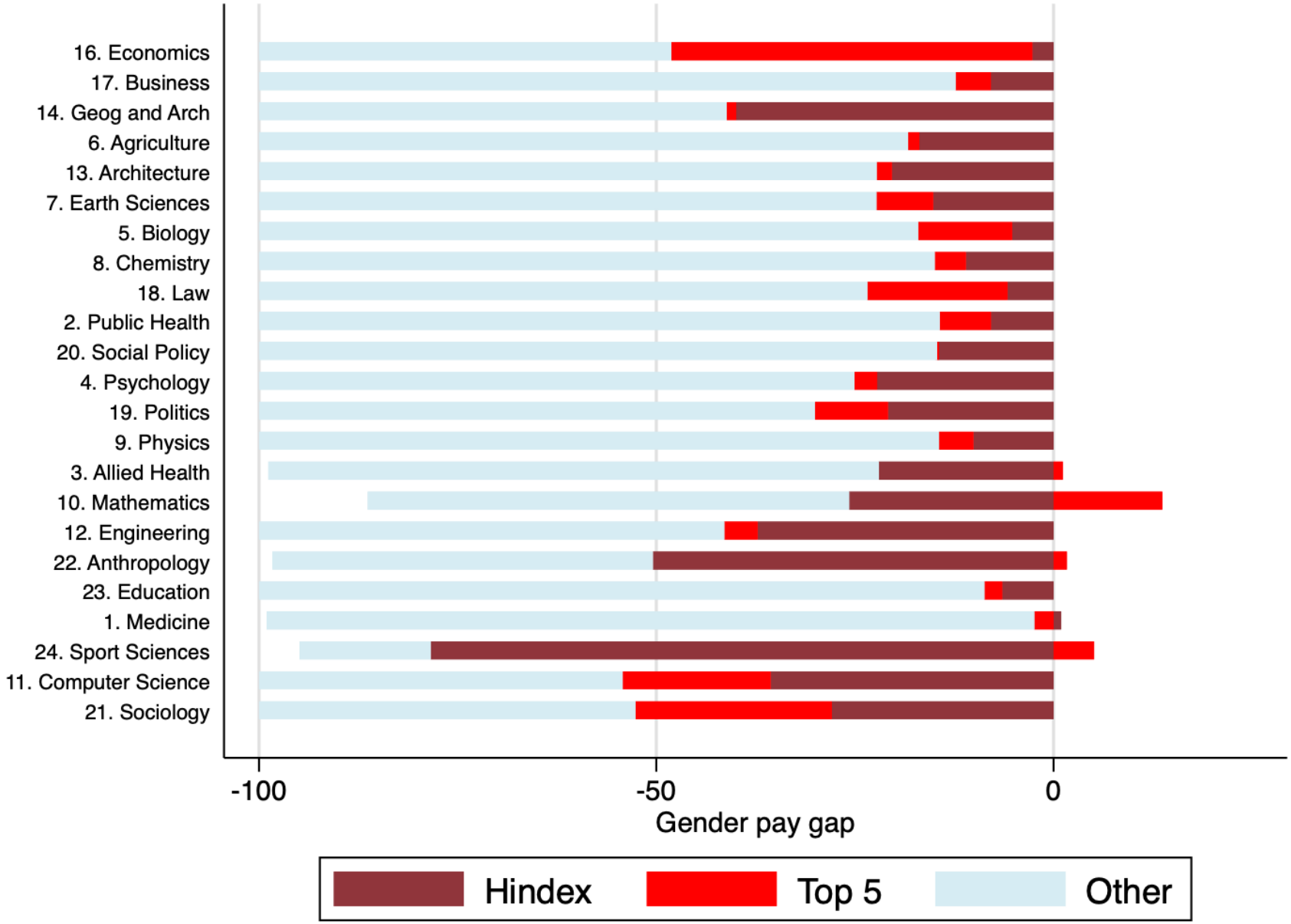
# Contribution to Gender Pay Gap in Economics

| | | | |
|---|---|---|---|
| Pay gap | $\overline{log(Y)}^F - \overline{log(Y)}^M$ | -0.178 | |
| *of which*: | | | |
| Outputs | | -0.053 | 31.5% |
| | *of which*: | | |
| *H-index* | $\widehat{\beta^H}\left(\overline{ln(H)}^F - \overline{ln(H)}^M\right) =$ | *-0.003* | *1.8%* |
| *Top 5* | $\widehat{\beta^{T5}}\left(\overline{T5}^F - \overline{T5}^M\right) =$ | *-0.053* | *29.7%* |
| Age (experience) | $\widehat{\beta^A}\left(\overline{A}^F - \overline{A}^M\right) + \widehat{\beta^{A2}}\left(\overline{(A^2)}^F - \overline{(A^2)}^M\right)$ | -0.061 | 34.5% |
| Unexplained | $\widehat{\beta^F}$ | -0.060 | 34.0% |

# Contributors to the pay gap



# Contributors to the pay gap, as % of total



# Discussion

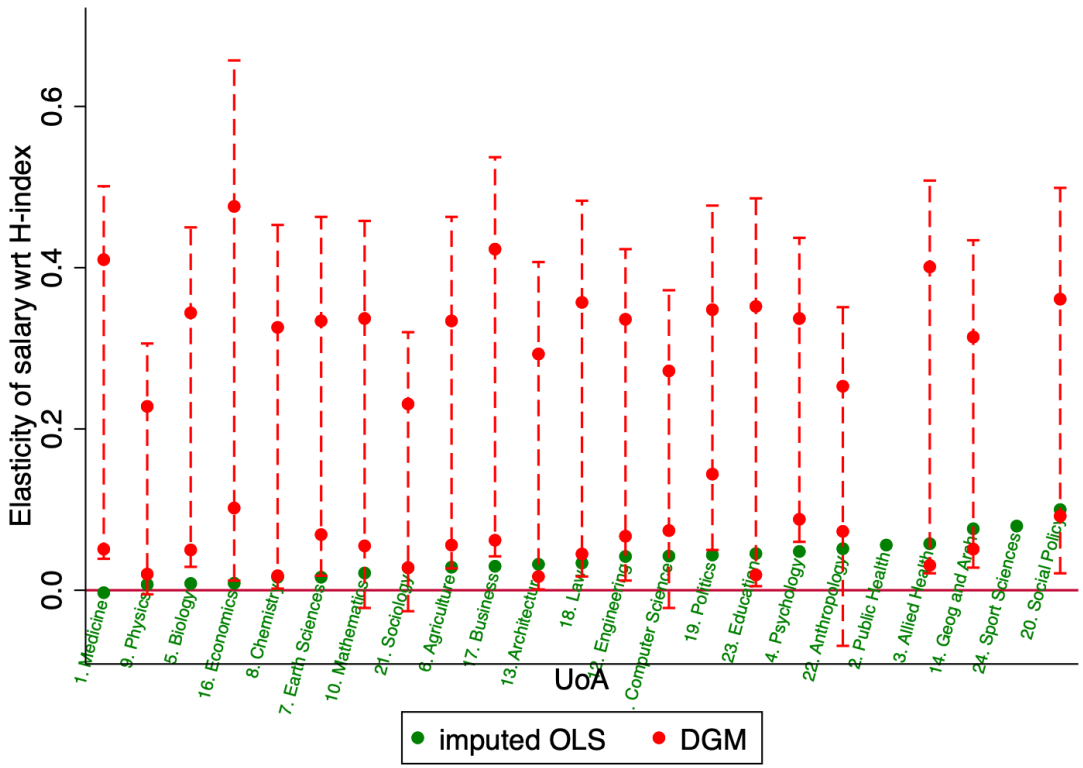
- Differences in outputs explain around a quarter of pay gap, more in some disciplines than others
- Economics differs from other disciplines in the
  - confluence of high surplus, high outside options, and publication norms that give high rewards to “big” papers that require large investments
  - the combination of differences in output levels and a high elasticity of salary wrt output explains a substantial portion of the gender pay gap
    - gender differences in Top 5 publications are higher in Maths and Chemistry the elasticity of salary wrt top 5 is low in these disciplines
    - the elasticity of salary wrt Top 5 is also sizeable in Computer Science, but the difference in number of Top 5 is small
- However, outputs do not explain the whole gender pay gap
- More work to do to estimate robust bounds on elasticities, include other covariates and utilising the panel

# Approaches to learn about $\beta$

1. Impute a value of $X$ into $Y$ data
   - nearest neighbour, sample, ...
2. DGM: D'Haultfoeuille, Gaillac and Maurel (2024) "Partially linear models under data combination" REStudies
   - $\beta = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$, we can get bounds on $\rho_{XY}$ using observed $\rho_{XW}, \rho_{WY}$
   - DGM provides a method to impose constraints, such as $\beta_{YX} \geqslant 0$, and on $R^2$, which help tighten the bounds
3. Manski and Tamer (2002) bounds
   - allows us to exploit that we have the population in both X and Y data
   - for each observed value of X what is the min and max of Y, what $\beta$ are consistent with that range

# DGM bounds on elasticity of salary wrt H-index

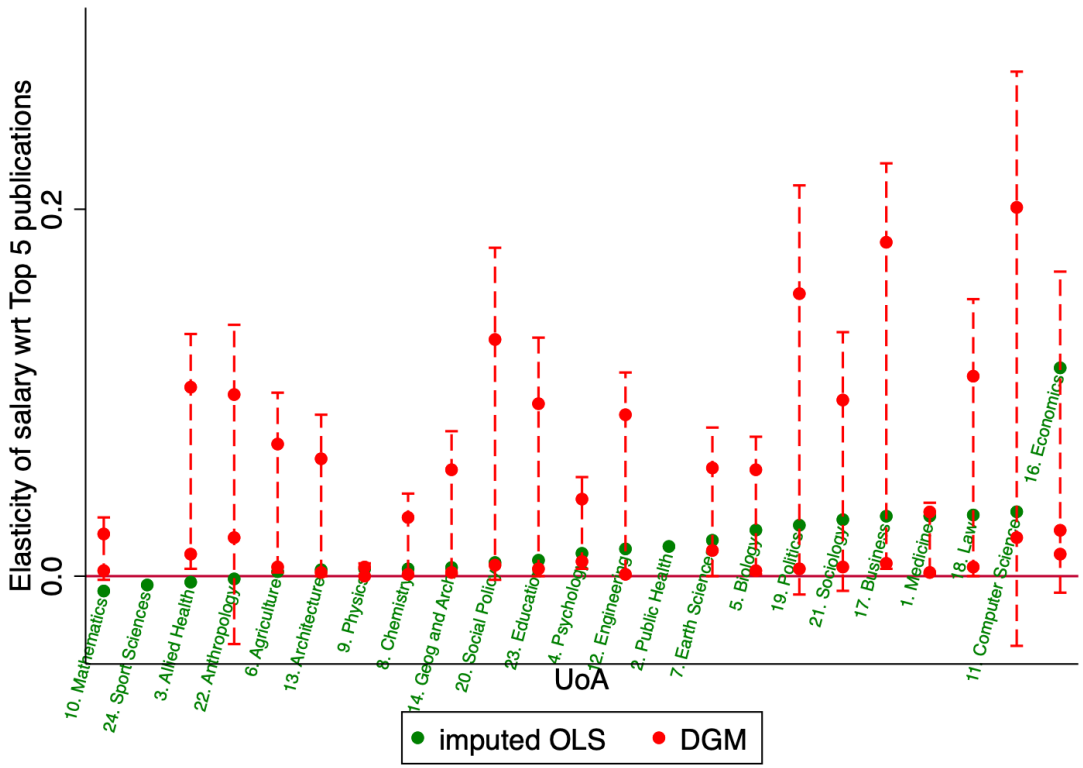
imposing $\beta^H \geqslant 0$, $\beta^F \leqslant 0$, $R^2 > \lambda R_I^2$




"imputed OLS" refers to estimates we get with OLS using imputed data.
"DGM" the dash lines show the 95% confidence interval obtained in table above, the two dots are lower and upper bounds of the set.

# DGM bounds on elasticity of salary wrt Top 5

imposing $\beta^H \geqslant 0$, $\beta^F \leqslant 0$, $R^2 > \lambda R_I^2$

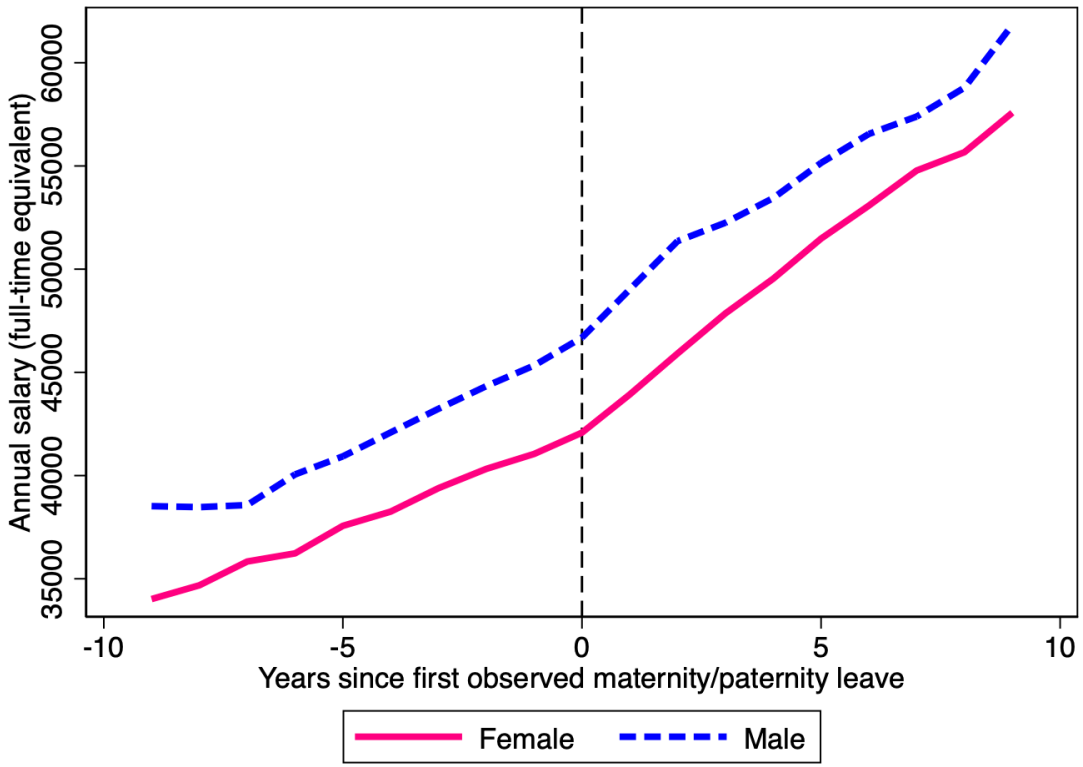



"imputed OLS" refers to estimates we get with OLS using imputed data.
"DGM" the dash lines show the 95% confidence interval obtained in table above, the two dots are lower and upper bounds of the set.

# EXTRA SLIDES

# Parental leave

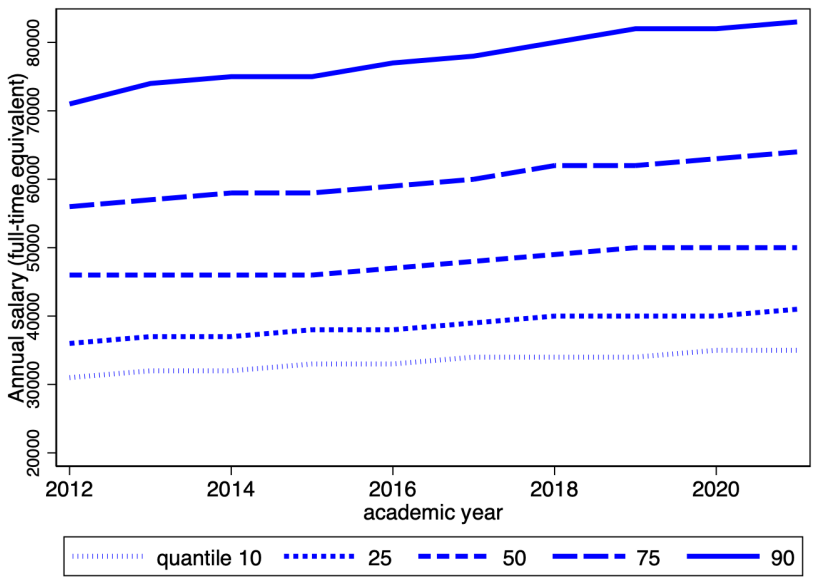
## research staff aged 29-40

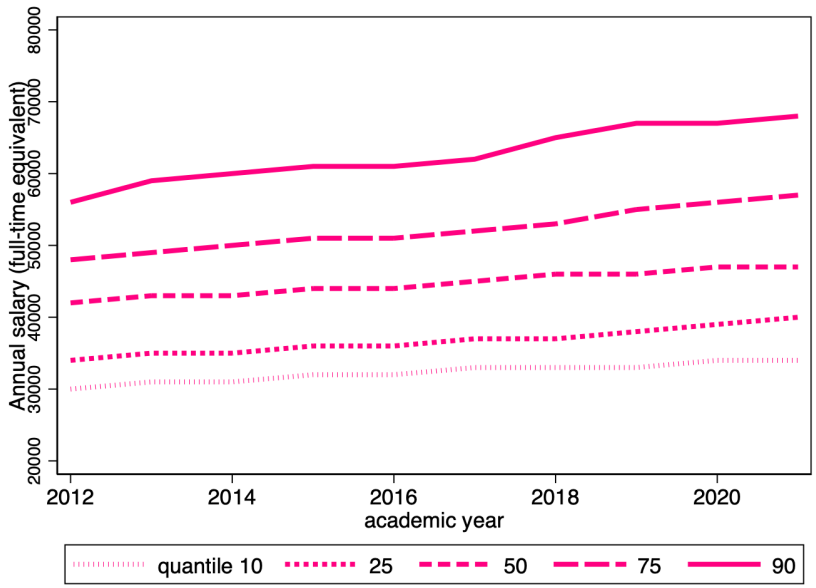



- 5% of females aged 31-40 take maternity leave (1.4% aged 21-30, 0.7% aged 41-50)
- 0.8% of males aged 31-40 take parental leave (0.2% aged 21-30, 0.3% aged 41-50)

# Male pay is higher than female at all quantiles

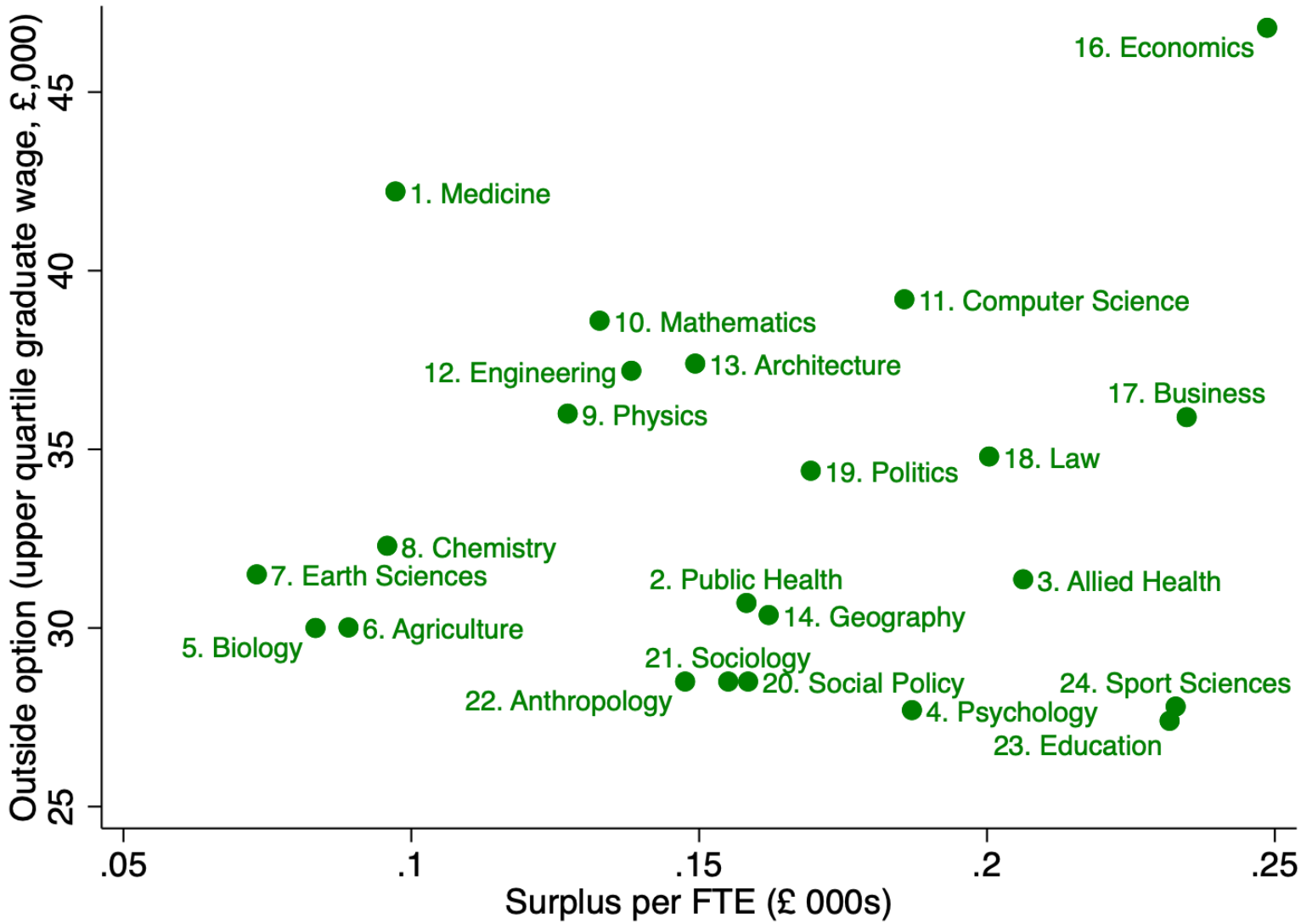
## research staff all universities




Source: authors' calculations using HESA data 2012-2021

# Surplus per FTE and Outside option

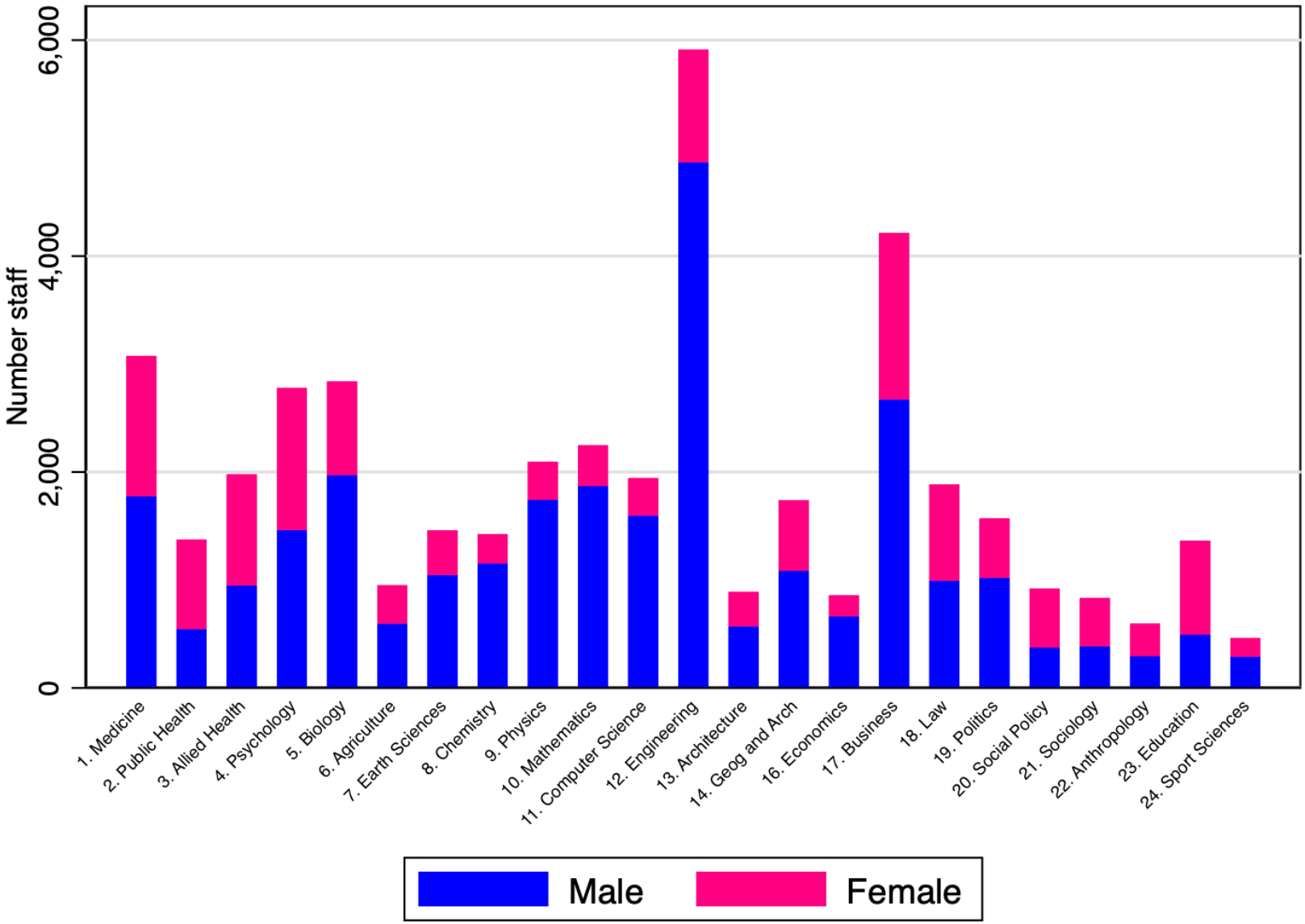



Source: HESA and LEO

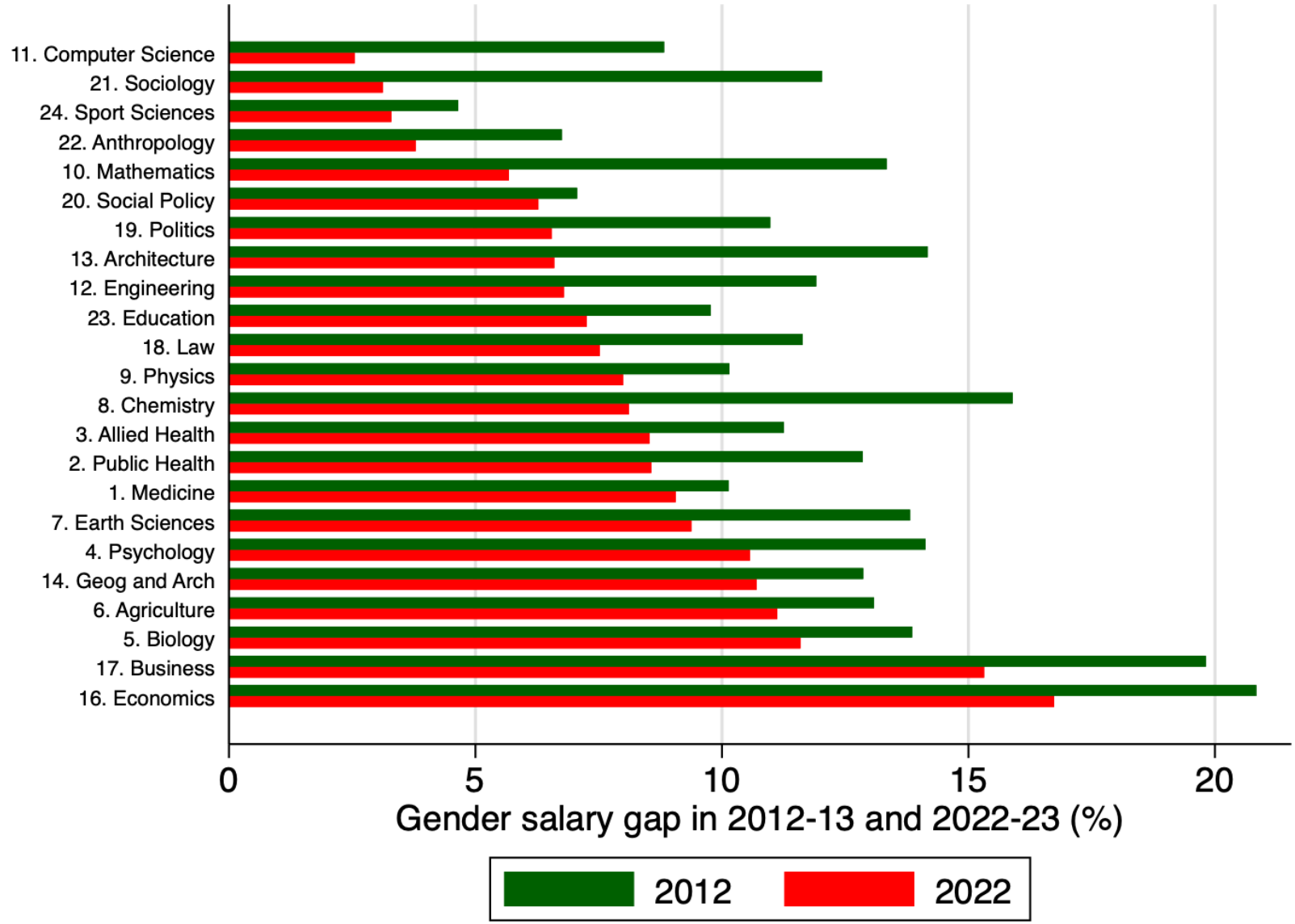
# Share of female academic staff

- All academic staff
  - in 2022-23 around 49% of staff are female, slightly up from 45% in 2012-13
  - this increase is almost entirely due to increase in staff on teaching contracts (from 27% to 37% of academic staff)
  - within teaching share of females has declined slightly from 48% to 46%
  - within research increased from 40% to 42%
- Staff submitted to REF2021
  - 33% were female
  - less than 25% in physics, maths, engineering, economics, to more than 50% in public health, social policy and education
  - share similar in REF2014 and REF2021

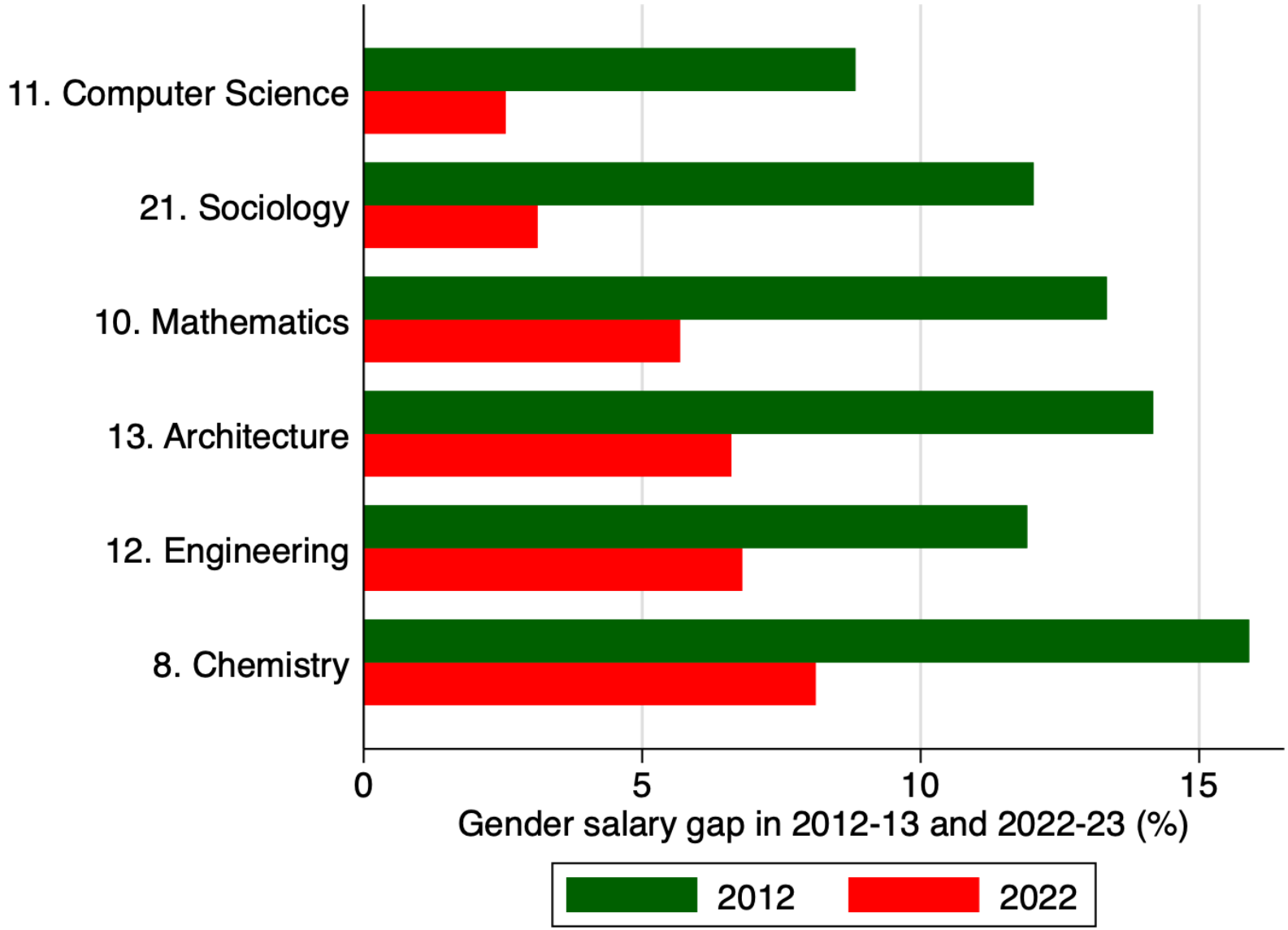
# Number of males and females



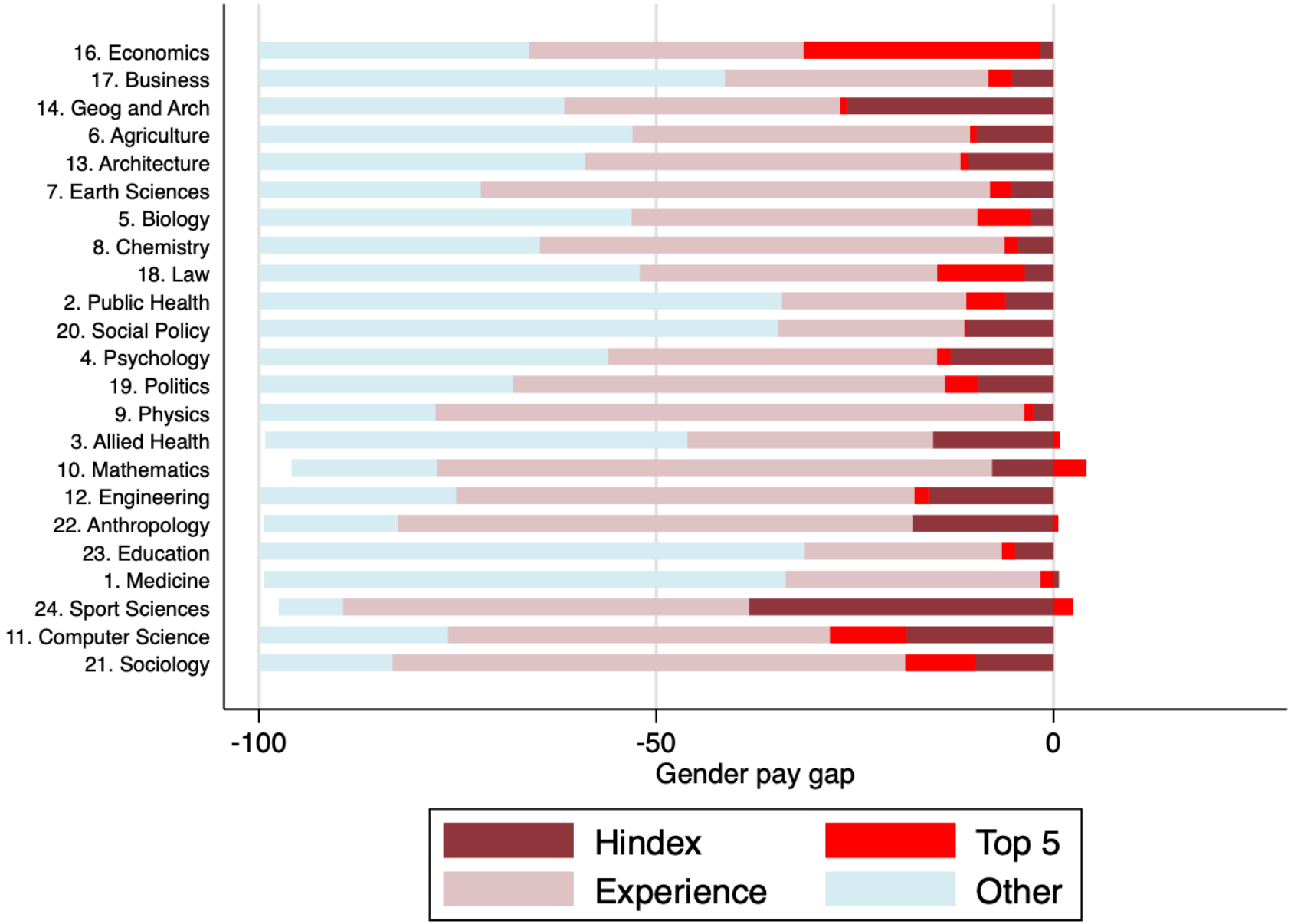
# Gender pay gap 2012 and 2022



# Gender pay gap sizeable reductions



# Contributors to the pay gap, as % of total



# We focus on research active staff

- We include all institutions with research income $>$15% of income
  - 56 institutions
  - 58% of staff, 66% of research active staff
  - 73% of outputs submitted to REF2021
- We exclude teaching focused institutions and specialist music, arts and agricultural institutions

# We focus on disciplines that publish in journals

23 disciplines

- Medicine, Health, Life Sciences (REF Panel A: UoA 1-6)
  - Medicine, Public Health, Allied Health Professions, Psychology, Biology, Agriculture
- Physical Sciences, Engineering, Maths (REF Panel B: UoA 7-13)
  - Earth Sciences, Chemistry, Physics, Mathematics, Computer Science, Engineering, Architecture
- Social Sciences (REF Panel C: UoA 14-24)
  - Geography, Archaeology, Economics, Business, Law, Politics, Social Policy, Sociology, Anthropology, Education, Sport Sciences
- ~~Panel D: Arts and Humanities (UoA 25-34)~~
  - ~~Area Studies, Modern Languages, English Literature, History, Classics, Philosophy, Theology, Art and Design, Performing Arts, Media Studies~~