Sorting and social mobility: an empirical matching model in higher education

Jack Britton (York and IFS)
Monica Costa Dias (Bristol and IFS)
David Goll (IFS)

Workshop in Labour Economics
Bank of Portugal and Bank of Spain
Tavira, June 2022

Motivation

- HE often seen as a key lever for social mobility
- The many policies incentivising disadvantaged students to enrol are meant to facilitate that process
- But though the returns to HE are large, they are also highly heterogeneous
  - ⋆ by the characteristics of HE programmes
  - ⋆ those of students
  - ⋆ and how the two align
- Those from poorer backgrounds benefit less from the HE system (Black et al. 2015, Campbell et al. 2019, Chetty et al. 2020)
- Here we investigate why and what policy can do about it

Questions

1. How do students sort into HE programmes in England?

   → In particular, to what extent is sorting determined by the abilities of students and their background, the characteristics of programmes and the returns to the investment?

2. In turn, what explains the variation in the labour market outcomes of University graduates?

3. To what extent is sorting into HE keeping low SES children from benefiting?

4. How does HE policy affect sorting and the distribution of returns?

What we do

Document key features of the HE market in England

Build structural model of life-cycle labour supply, earnings and education with an embedded equilibrium model of the HE market

→ Separates preferences of students and programmes for each other

Estimates based only on demand side fail to account for (unobserved) choice sets

We consider that observed allocation reveals a constrained choice, where the choice sets are unobserved

Exploit instrumental variation and structure of market for identification of preferences and returns to HE

Use model to understand implications of implemented reforms and for counterfactual analysis

Key findings

1. Equilibrium model of HE market fits data well and predicts well reduced form estimates of the impact of tuition and loan reforms
2. Strong assortativeness between students' skills and programmes' characteristics
3. Evidence of earnings complementarities between student's and programme characteristics
4. Programme quality (akin to selectivity) has small but significant impact on earnings; field of study matters more, as do the abilities of students
5. Conditional on sorting, background still matters a lot
6. Demand-side policies have small impacts on the sorting of students and social mobility
7. Supply-side policies (quotas) can be more effective, but at a cost

Background

Returns to university quality and field of study

Dale & Krueger 2002, 2014, Back & Smith 2006, Broecke 2012, Hastings et al. 2013, Kierkeboen et al 2017, Anelli 2018, Dillon & Smith 2020

High education choices

Keane and Wolpin 1997, 2001, Arcidiacono 2004, Wiswall and Zafar 2015, Delavande and Zafar 2019

Modelling HE market

Gale Shapley (1969), Roth and Sottomayor (1992), Arcidiacono 2005, Epple et al 2006, Fu 2014, Kapor 2020

Identification in many-to-one matching markets

Agarval 2015, Agarval and Diamond 2017

The effects on grant, tuition and loan policies

Denning et al 2019, Epple et al. 2006, Hubner 2012, Azmat and Simion 2020

Institutional features: the HE market

Students apply in their final HS year, *before sitting final exams* (A-levels)

- ⋆ Choose up to 5 programmes (subject×university combinations)
- ⋆ Demonstrate ability by test scores at age 16 (GCSEs), University interviews and tests

The University system is public

- ⋆ Universities have a set number of places
- ⋆ Charge tuitions that are regulated by law

Sorting

- ⋆ Universities make offers (conditional on final results)
- ⋆ Once A-level results known: students take preferred standing offer
- ⋆ ... or go through 'clearing', where remaining places are allocated

HE funding policy

2006-2011

- Tuition fees capped at £3,000 per year, apply to every student
- No upfront costs: student loans cover tuition + living costs (up to cap)
- Repayment: 9% of income above threshold (£15,000 in 2011)
- Interest rate averaged 1.5%
- Outstanding debt forgiven after 25 years

Large reform in 2012

- Big fee increases: trebled to £9,000
- Interest rates increased to RPI + 3%
- Repayment threshold raised (£15,000 to £21,000)
- Loan term extended (25 to 30 years)

Data

Longitudinal Educational Outcomes (LEO)

English administrative data, links information from three sources:

- **National Pupil Database (NPD)**
  School id, test scores at ages 11 (SATs), 16 (GCSEs) and 18 (A-levels), gender, cohort, ethnicity, place of residence, local index of deprivation (IDACI), free school meals
- **Higher Education Statistics Agency (HESA)**
  University and field of study for those who enrol in HE
- **HMRC tax records**
  Working status and annual earnings

Study cohorts born in 1988-91, entering university in 2006-09

Over 2.5 million children, 35% enrol in 3-year University programmes

Heterogeneity in HE programmes

- By subject

  earnings by subject

  LEM: Law, Economics and Management

  STEM: Science, Engineering, Maths and Medicine

  Soc Sc: Arts, Humanities and Social Sciences

- By University: all 150 institutes in England

  earnings by institution

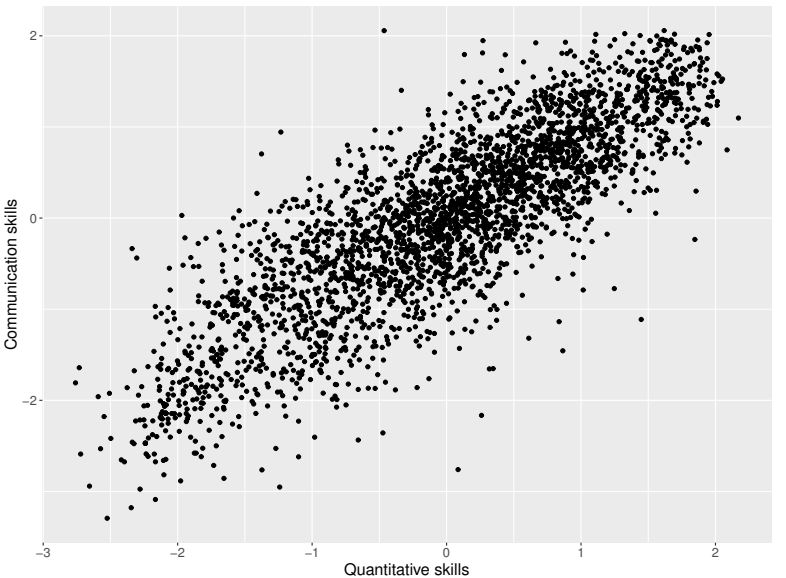
- Summary measure of **programme quality**
  - ⋆ *Complete University Guide* – league tables for all UK programmes
  - ⋆ First principal component, 5 quality indicators: spending in academic services and facilities, research quality, student/faculty ratio, student satisfaction

Skills of students

- Two dimension of skills: communication and quantitative
- Use factor analysis to summarise test scores at 11 and 16
- Exclusion restrictions: Maths and English scores at 16

Skills of students

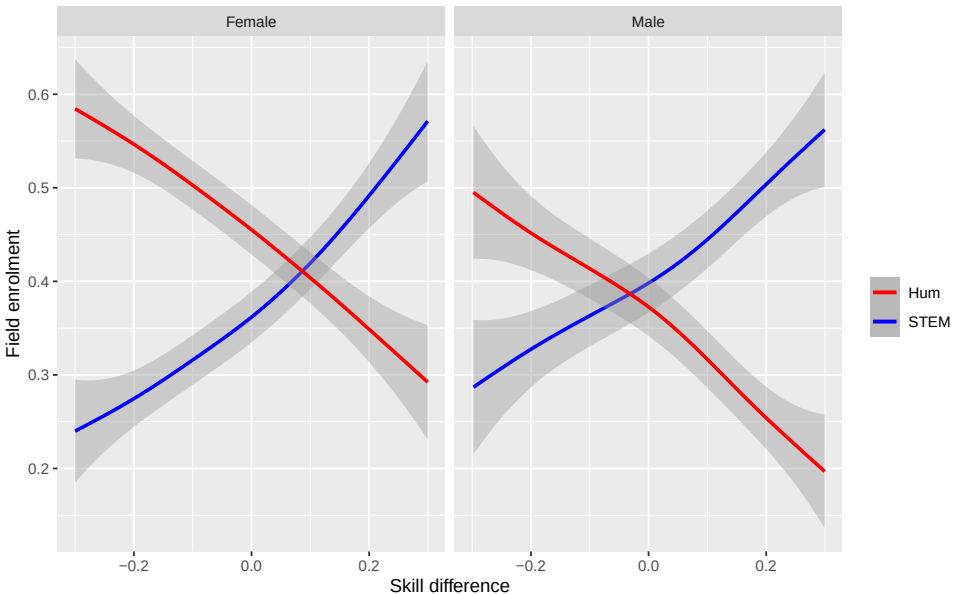
- Two dimension of skills: communication and quantitative
- Use factor analysis to summarise test scores at 11 and 16
- Exclusion restrictions: Maths and English scores at 16



Communication and quantitative skills (corr=0.9)

High Education Sorting

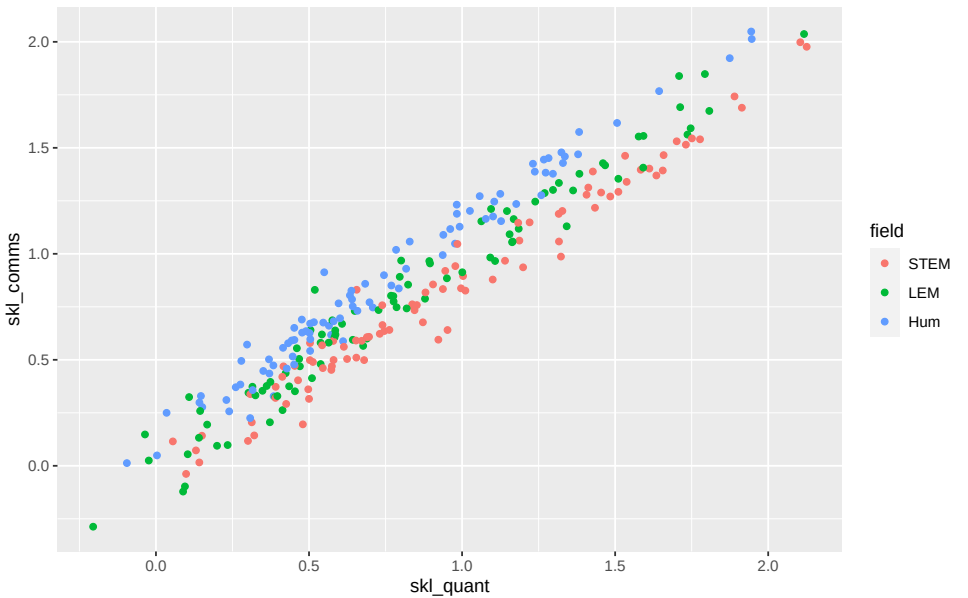
Relative skills matter for subject choice



Relative skills and subject sorting

High Education Sorting

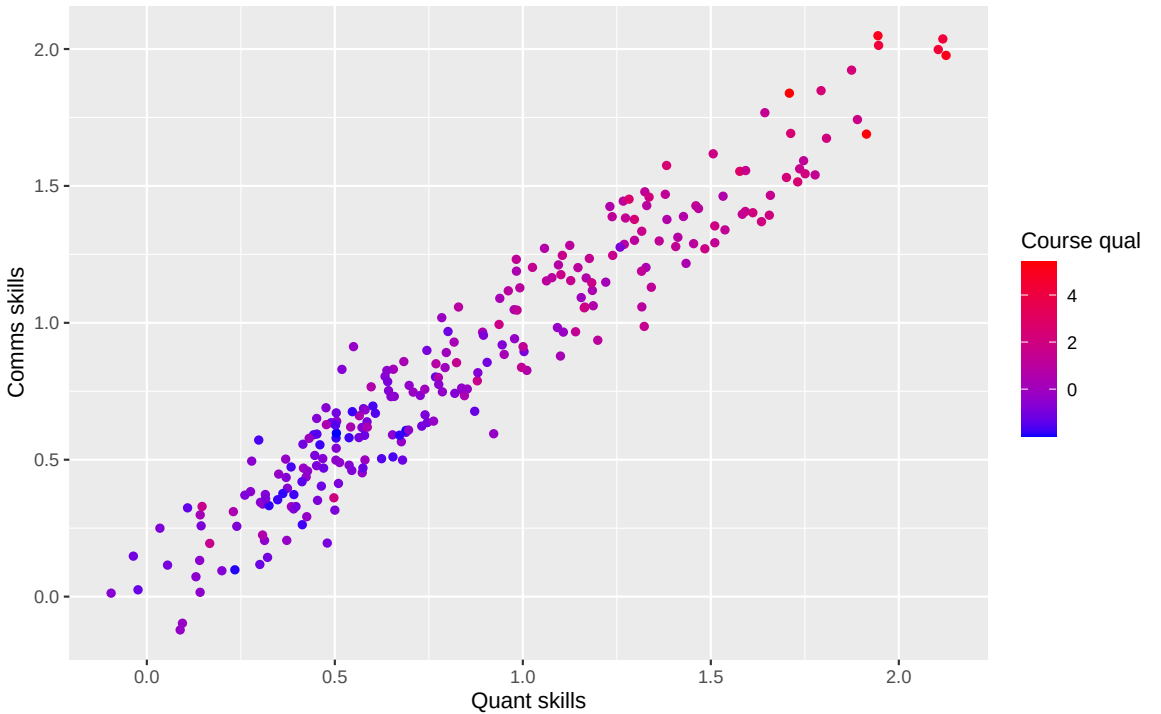
And they matter across the distribution



Average skill composition by course

High Education Sorting

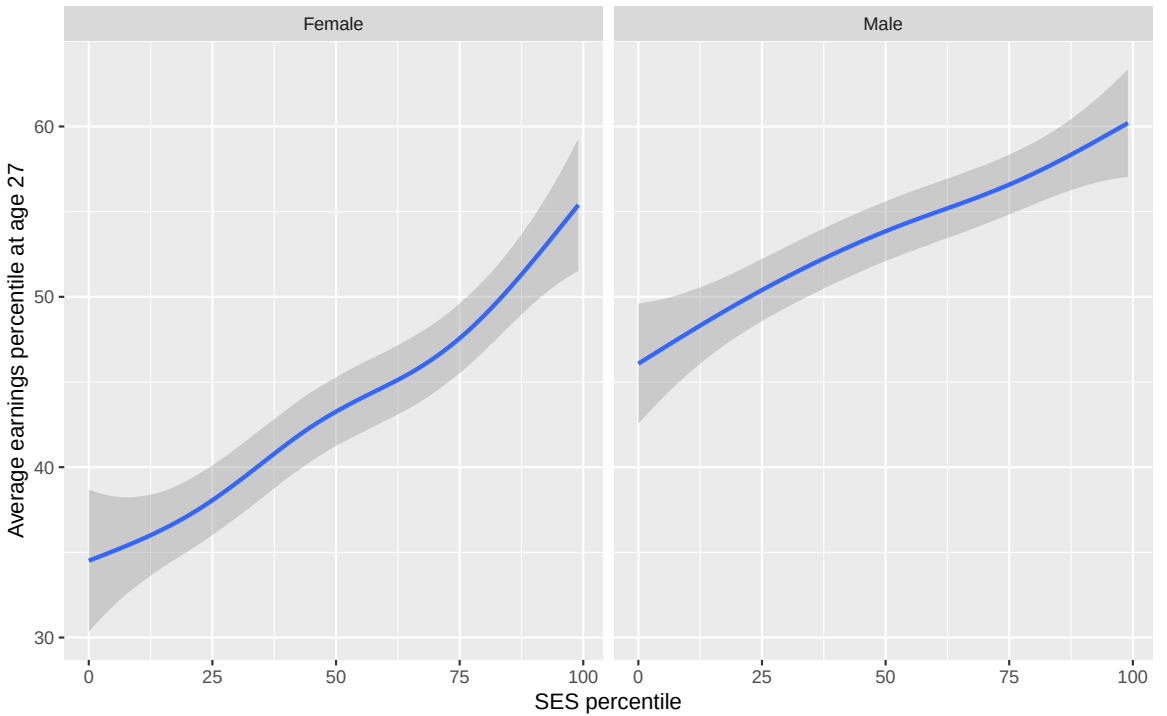
While the quality index captures programme selectivity



Average skill composition by course

Socio-economic gradient

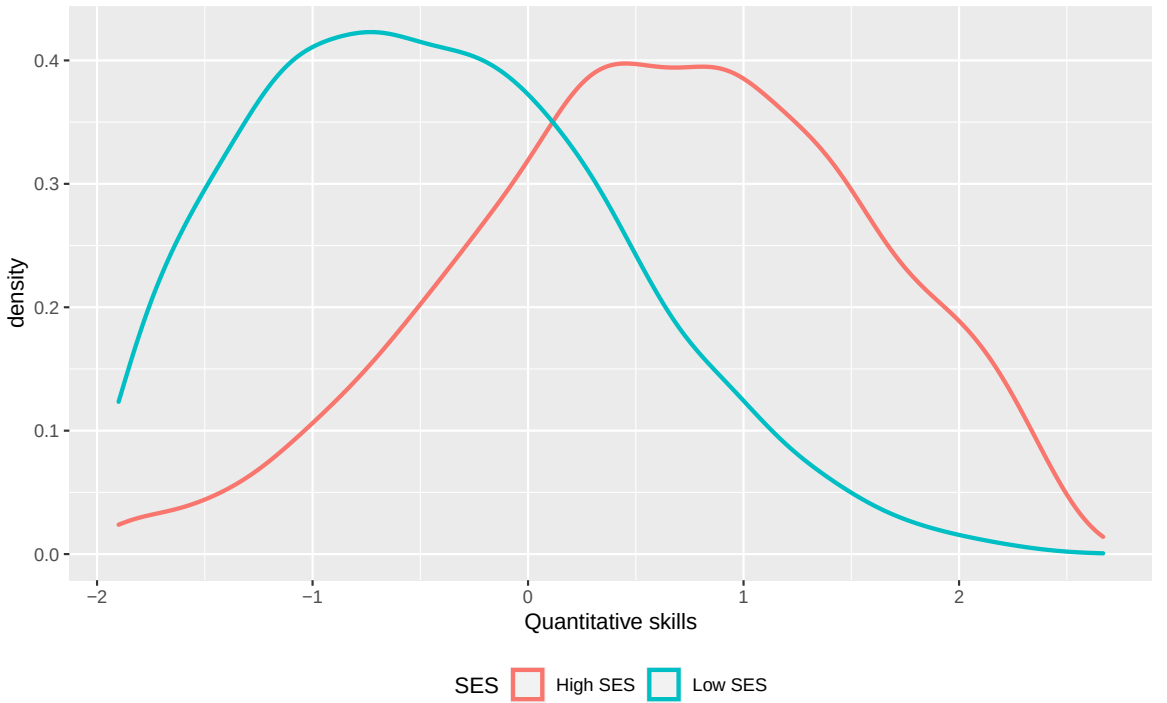
Background disadvantage strongly reflected on later earnings



Parental SES and child's income

Socio-economic gradient

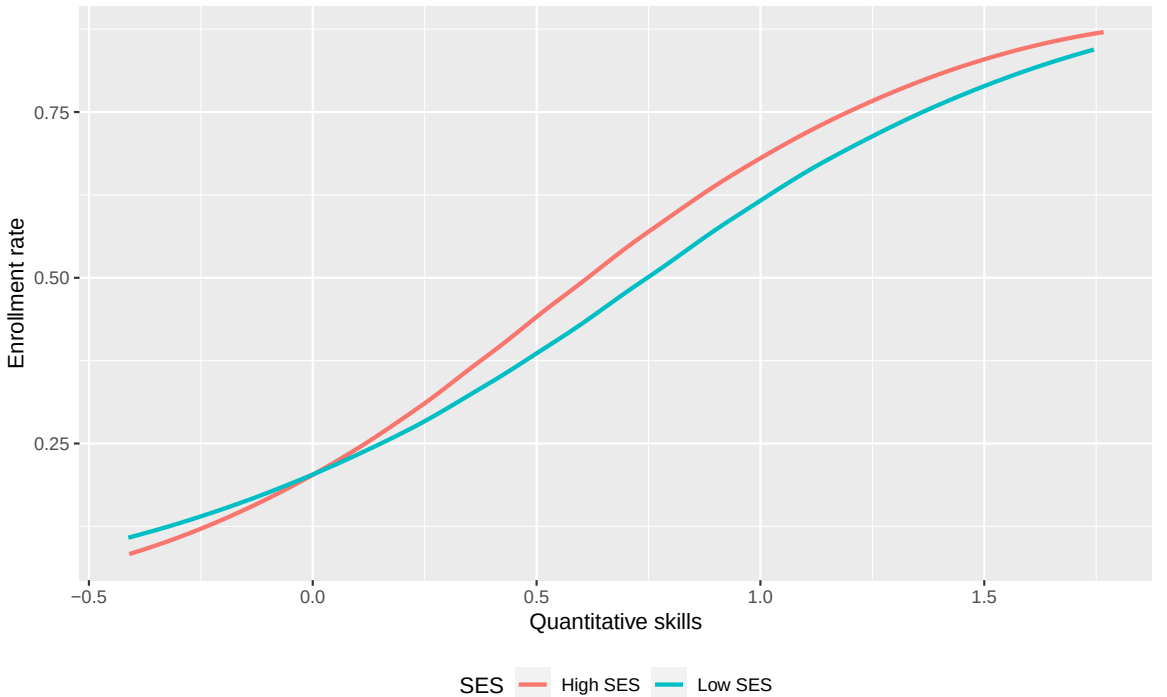
... but also on skills



Distribution of quantitative skills by parental SES

Socio-economic gradient

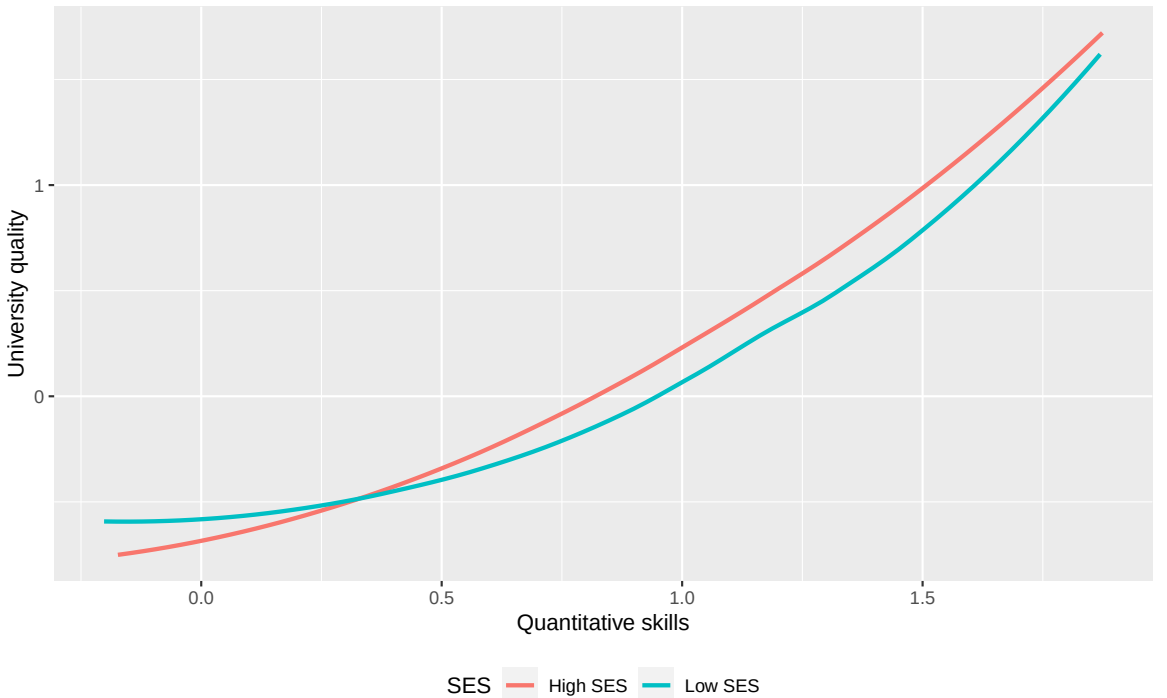
... participation in HE



Enrollment rates by quantitative skills and parental SES

Socio-economic gradient

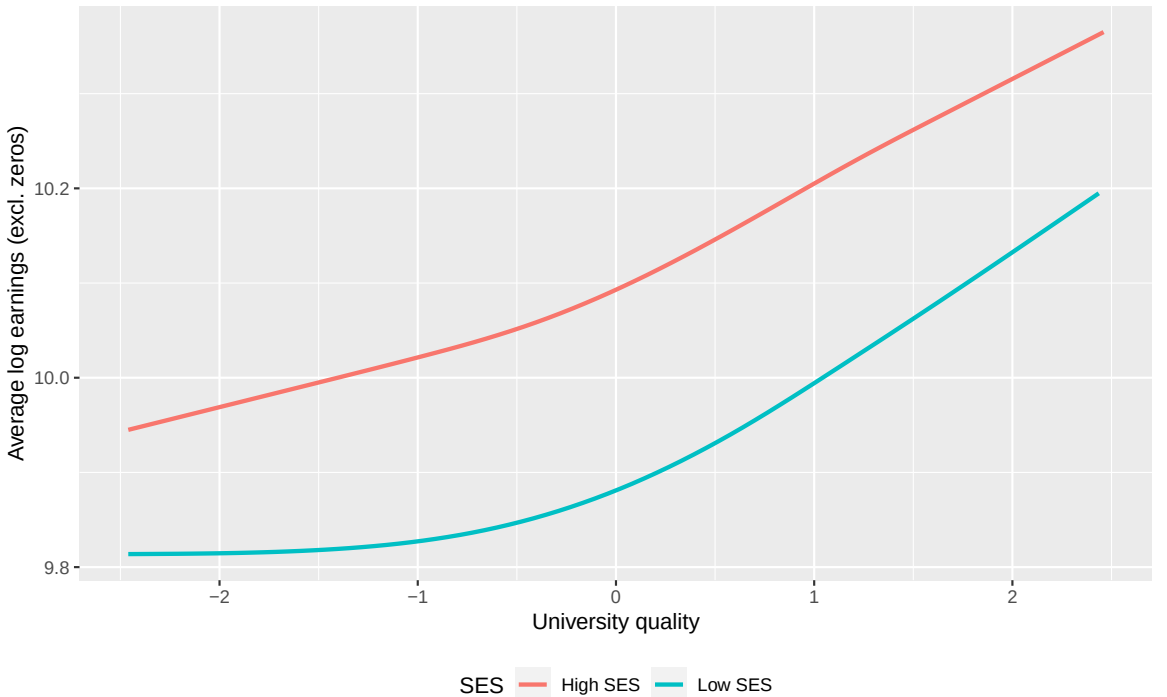
... the selectivity of degree



Average programme quality among participants, by quantitative skills by parental SES

Socio-economic gradient

... and earnings conditional on selectivity



Average earnings by programme quality and parental SES

Model

Life in two stages

Stage 1: education

- ⋆ Heterogeneous students decide whether to participate in HE
- ⋆ Students and programmes meet in many-to-one matching market without frictions
- ⋆ Tuition fees set exogenously, uniform across programmes
- ⋆ Preferences and capacity constraints determine equilibrium sorting
- ⋆ Non-transferable utility

Stage 2: work

- ⋆ Simplest life-cycle model of labour supply, consumption and savings
- ⋆ Earnings depend on education attainment and characteristics of students and the programmes they attended

Model

HE programmes rank applicants

- Heterogeneous programmes $j \in J$ offer education valued in work
  - ⋆ by subject F_j: LEM, STEM and Soc Sc
  - ⋆ and quality Q_j
- Care about reputation, reflected in skill composition of their students:
  - ⋆ observed mathematical and communication skills (S^m, S^c)
  - ⋆ unobserved (to the econometrician) productive ability θ
- Strict preference for filling available spots

Subject-specific preferences for student i:

$$W_{ji} = \gamma^X X_i + \gamma^m_{F_j} S^m_i + \gamma^c_{F_j} S^c_i + \gamma^\theta_{F_j} \theta_i + \epsilon_{F_j,i}$$

ϵ: unobserved preferences – skills that do not matter for earnings or errors on the part of programmes in assessing student's skills

Model

Students rank programmes

- Students $i \in I$ characterised by
  - ⋆ Skills (S_i^m, S_i^c, θ_i)
  - ⋆ Traits X_i: socio-economic background $(\text{SES}_i, \text{private}_i)$, gender g_i
- Draw value from attending HE and its returns
  - ⋆ Varies with nature of the programme (F, Q)
  - ⋆ ... distance from home – distance_{ij}
  - ⋆ ... choices of past cohorts in secondary school – share_{ij}

Value for student i of joining programme j

$$U_{ij} = u\left(F_j, Q_j, \text{distance}_{ij}, \text{share}_{ij}, S_i^m, S_i^c, X_i\right) + \eta_{ij} + \text{E}V_{ij}$$

η: unobserved preferences for programme j, two components – preferences for subject and university

Model

Earnings

Pre-tax earnings t years out of education with qualification j in field f

$$
\begin{aligned}
\ln(y_{it}) &= \alpha_{0f} + \alpha_{1f} X_i + \alpha_{2f} S_i^m + \alpha_{3f} S_i^c + \alpha_{4f} Q_j + \\
&\quad + \alpha_{5f} Q_j S_i^m + \alpha_{6f} Q_j S_i^c + \alpha_{7f} \ln(t+1) + \alpha_{\theta f} \theta_i + \zeta_{fit} + \iota_{it} \\
\zeta_{fit} &= \rho_f \zeta_{fit-1} + \xi_{fit} \\
\iota &\sim iid
\end{aligned}
$$

Model

EV is solution to simple life-cycle problem

- University students use their loans to consume and pay fees
- Working life starts at 22, or 19 for non-graduates

$$
\begin{aligned}
V_{it}^{j} &= \max_{c_{it},d_{it},a_{it+1}} \left[\ln(c_{it}) + g\left(t, g_i, f\right) d_{it} + \beta \mathrm{E}V_{it+1}^{j}\right] \\
\text{st} \quad a_{it+1} &= R a_{it} + d_{it} y_{it} - P\left(d_{it} y_{it}, l_{it}\right) - T\left(d_{it} y_{it} - P\right) - c_{it} \\
l_{it+1} &= R_l l_{it} - P\left(d_{it} y_{it}, l_{it}\right) \\
l_{it} &\geq 0
\end{aligned}
$$

(R, R_l): risk-free and student loan re-payment interest rates

P: repayment schedule

T: tax liabilities and benefit entitlements

Match equilibrium

- Centralised market with non-transferable utility, a la Gale-Shapley
- No prices in this market: programme selectivity rule (or capacity constraints) act as prices to clear market (Azevedo and Leshno 2016)
- A stable match exists (Roth and Sottomayor 1992, Abdulkadiroglu et al. 2015, Azevedo and Leshno 2016)
- We assume stability, which means

$$U_{ij} > U_{i,\mu(i)} \implies \min_{i' \in \mu^{-1}(j)} \{W_{ji'}\} > W_{ji}$$

  where
  - $\mu(i) : I \to J$ is the programme attended by student i
  - $\mu^{-1}(j)$ is the set of students matched to course j

Estimation

Simulated Method of Moments: estimate parameters Θ using moments M:

$$\hat{\Theta} = \min_{\Theta} (M - M(\Theta))' \mathbf{W} (M - M(\Theta))$$

Θ includes parameters in:

1. Student utility
2. Earnings
3. Course preferences

Data: Cohorts finishing high school in 2006-09 (pre-reform)

Two parameters are fixed: discount rate $\beta = 0.95$ and the t coefficient in the earnings equation, which is estimated outside the model

Identification

Challenges

1. Separate preferences of students from those of programmes
2. Identify returns to programmes in the labour market and distribution of unobserved ability

Identification I

Excluded variables from earnings and university preferences

Distance to university and **choices of past cohorts in school**

→ Vary demand for programmes without changing choice sets

→ Vary education outcome w/o affecting potential earnings estimates

∴ Helps separating returns to programmes from distribution of student types by programme

Identification II

Sorting patterns

Learn about how students (programmes) trade-off programme (student) characteristics
(Chiappori ??)

→ Suppose two programmes differ along 2 characteristics and are matched to a similar group of students

→ Means that these programmes are equally desirable to students

→ Reveals how students trade-off these programme characteristics

∴ Include moments on correlation between student & course characteristics

Identification III

Many-to-one matching

Separate programme and students valuations
(Agarwal 2015, Diamond 2017)

If 'within variation' << 'between variation' then programmes should have strong preference for characteristic

E.g. if students' quantitative skills are highly predictive of their academic performance, then universities will place a high weight on that characteristic and variation in that skill within courses should be small

$\rightarrow$ Helps identifying how universities value the characteristics of students

$\therefore$ Include moments describing within and between programme variation in student characteristics and correlations between one's characteristics and those of their peers

Parameter estimates

Earnings equation

$$\begin{aligned}\ln(y_{it}) &= \alpha_{0f} + \alpha_{1f}X_i + \alpha_{2f}S_i^m + \alpha_{3f}S_i^c + \alpha_{4f}Q_j + \\ &+\alpha_{5f}Q_jS_i^m + \alpha_{6f}Q_jS_i^c + \alpha_{7f}\ln(t+1) + \alpha_{\theta f}\theta_i + \epsilon_{fit}\end{aligned}$$

	No Univ	STEM	LEM	Soc Sc
intercept	9.90*	10.08*	10.10*	9.99*
female	-0.31*	-0.15*	-0.19*	-0.13*
SES	0.32*	0.20*	0.21*	0.23*
private school	0.05*	-0.03*	-0.01*	0.00
math skills	0.10*	0.09*	0.13*	0.11*
communication skills	0.05*	0.05*	0.03*	0.01*
programme quality		0.02*	0.03*	0.02*
quality × math		0.03*	0.01*	0.02*
quality × communication		-0.00	0.01*	0.01*
unobserved skill	0.04*	0.04*	0.04*	0.04*

* Significant at 5% level

Parameter estimates

Student preferences for programmes

$$U_{ij} = u\left(F_j, Q_j, \text{distance}_{ij}, \text{share}_{ij}, S_i^m, S_i^c, X_i\right) + \eta_{ij} + \text{EV}_{ij}$$

	STEM	LEM	Soc Sc
intercept	0.02*	-0.79*	1.83*
female	-0.05*	-0.80*	-0.04*
SES	0.80*	0.43*	0.78*
private school	0.38*	0.36*	0.37*
math skills	0.01*	-0.01*	-0.01*
communication skills	0.01*	0.00	-0.03*
school share	0.10*	0.21*	
distance		-0.52*	
distance×SES		0.74*	
distance×private		0.27*	
Variance: field preference		1.01*	
Variance: university preference		0.57*	

* Significant at 5% level

Parameter estimates

Programme preferences for students

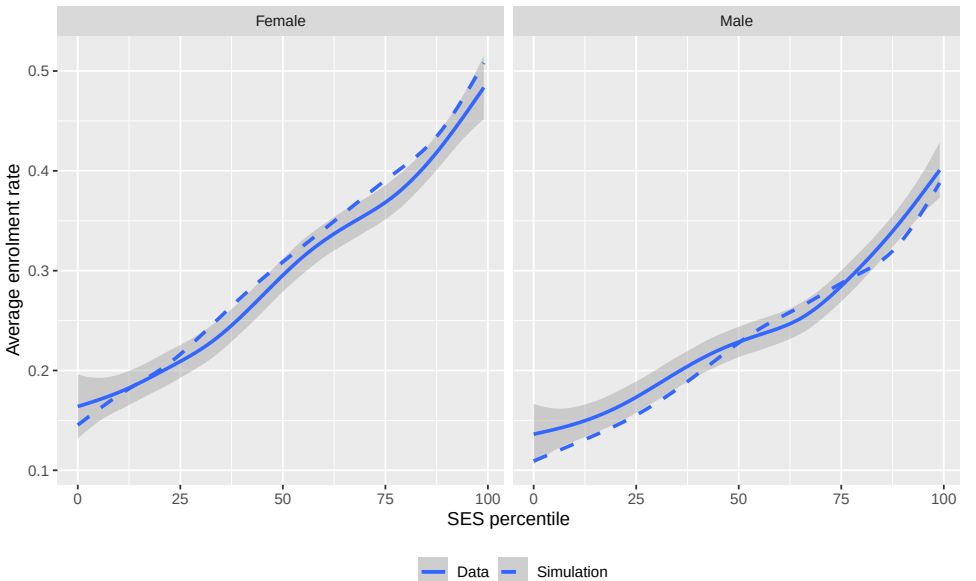
$$W_{ji} = \gamma^X X_i + \gamma^m_{F_j} S^m_i + \gamma^c_{F_j} S^c_i + \gamma^\theta_{F_j} \theta_i + \epsilon_{F_j,i}$$

	STEM	LEM	Soc Sc
female		-0.06*	
SES		0.02*	
private school		0.30*	
math skill	1.57*	0.61*	0.11*
communication skill	0.05*	0.82*	1.44*
unobserved skill		0.05*	

* Significant at 5% level

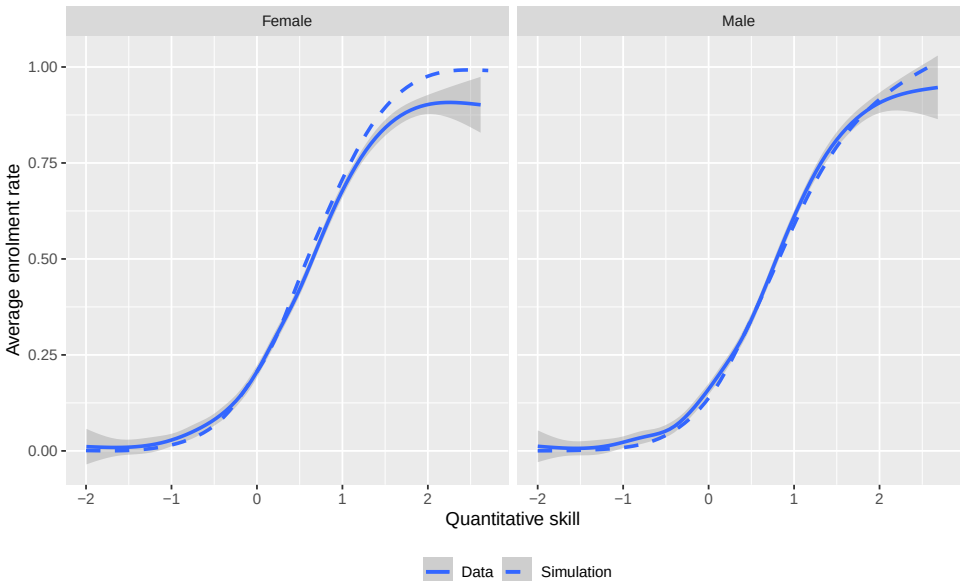
Model fit

Enrolment by SES and by gender



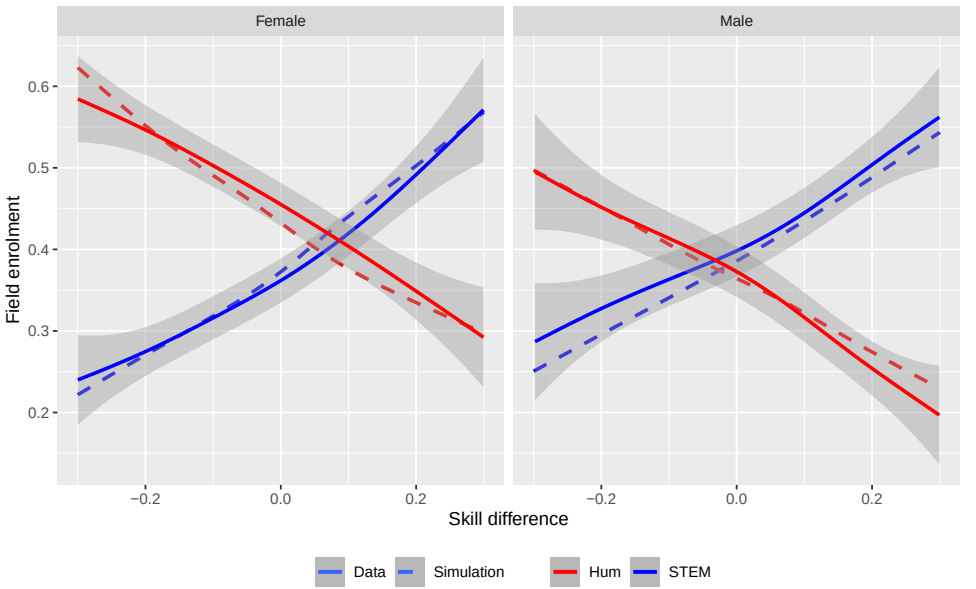
Model fit

Enrolment by quantitative skills, by gender



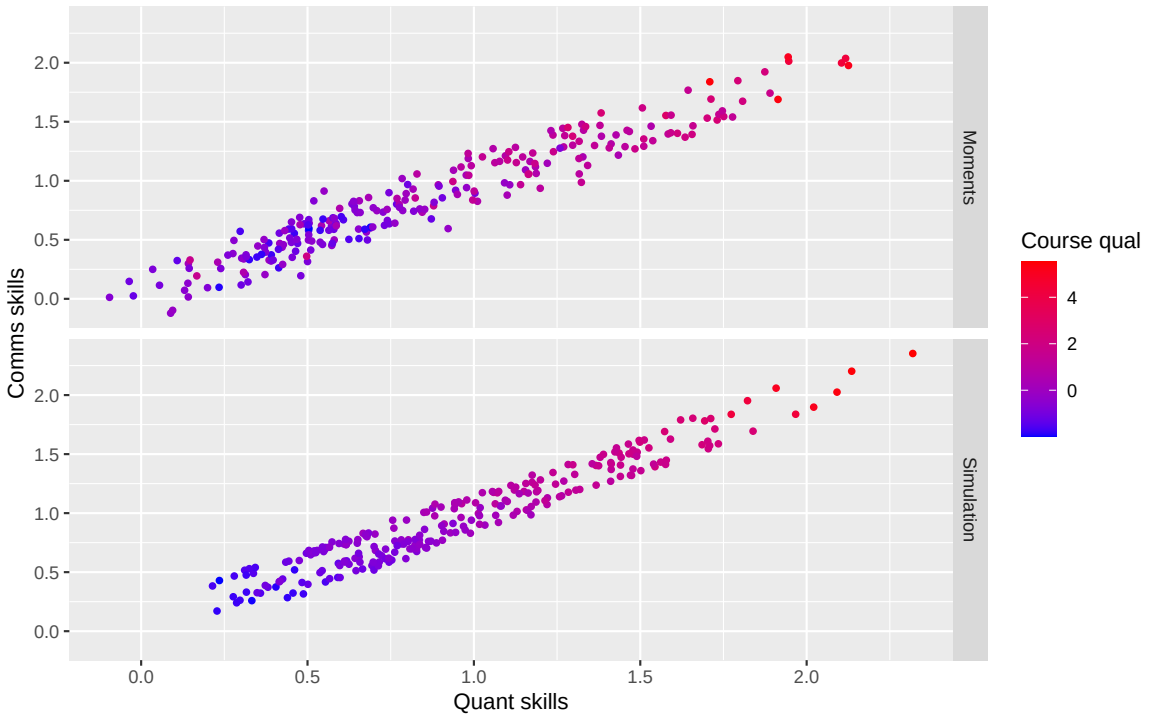
Model fit

Relative skills and subject choice, by gender



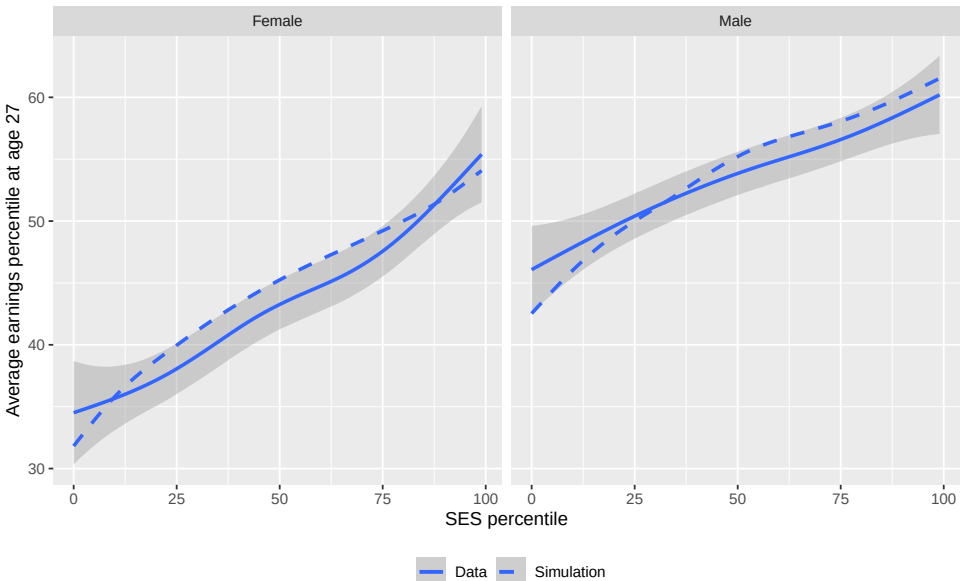
Model fit

Skill composition by quality of programme



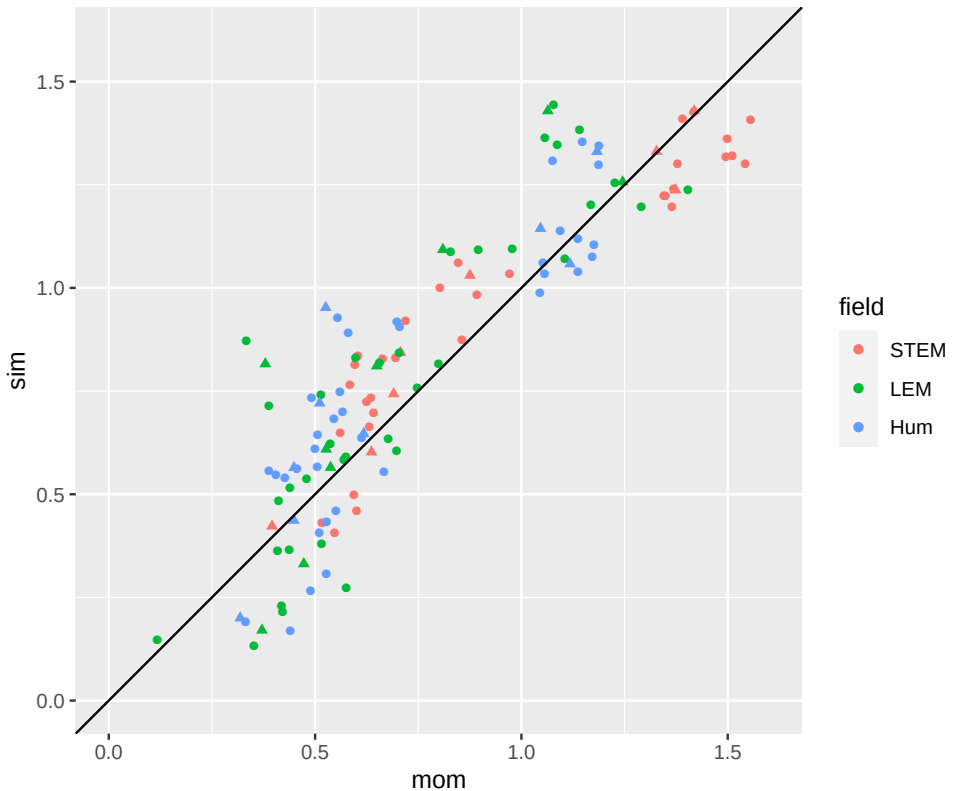
Model fit

Social mobility, by gender



Model fit: external validation

Average quantitative skills by programme, in and out of sample



HE funding policy

2006-2011

- Tuition fees capped at £3,000 per year in 2006-11, apply to every student
- No up-front costs: student loans cover tuition + living costs (up to a cap)
- Repayment: 9% of income above threshold (£15,000 in 2011)
- Interest rate averaged 1.5%
- Outstanding debt forgiven after 25 years

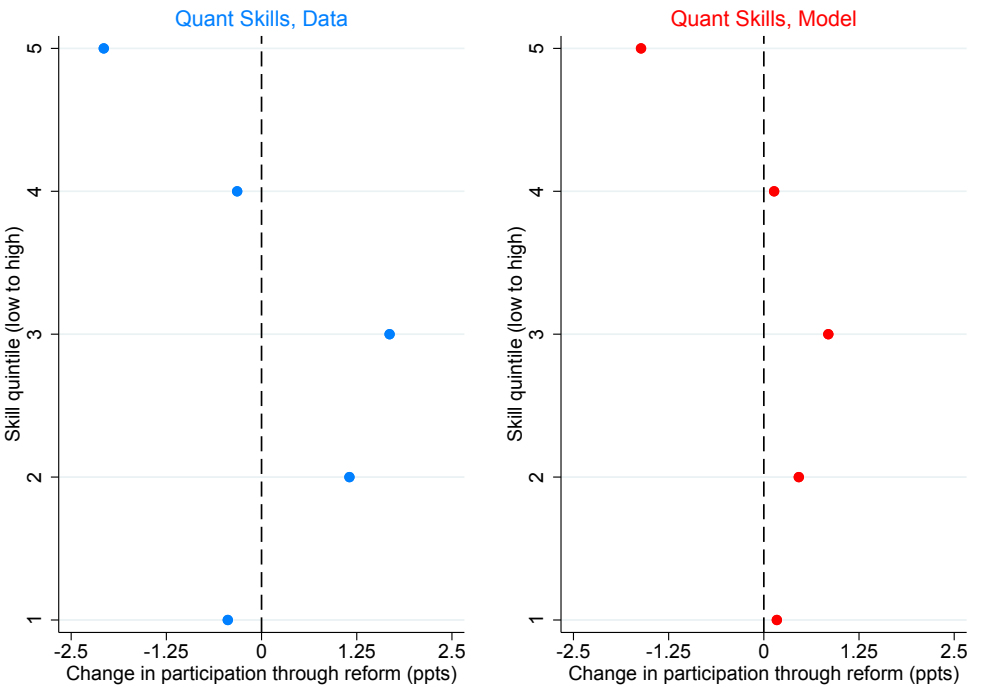
Large reform in 2012

- Big fee increases: trebled to £9,000
- Interest rates increased to RPI + 3%
- Repayment threshold raised (£15,000 to £21,000)
- Loan term extended (25 to 30 years)

repayment schedules

Model fit: external validation

Enrolment effects of 2012 reforms



Counterfactual policy reforms

- **Policy 1:** Additional maintenance grants of £4,000 per year, on top of existing loans, for low SES (bottom 50%) students
- **Policy 2:** As above, but only if studying STEM
- **Policy 3:** No student loans for low SES students - tuition fees set to £0 and maintenance loans converted to grants at current levels
- **Policy 4:** 10% rule giving preferential admission to low SES students only who graduated in top 10% in their school

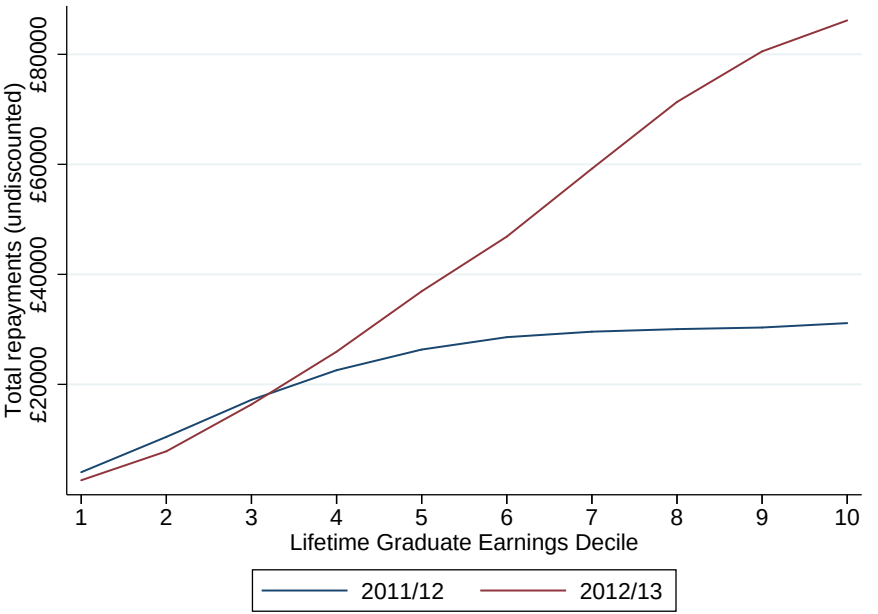
Counterfactual policy reforms

	Baseline	P1 Grants	P2 STEM Grants	P3 Loan write-off	P4 Cond 10% rule
			Low SES		
HE participation	0.20	0.20	0.20	0.20	0.22
STEM share	0.45	0.44	0.55	0.45	0.45
Undermatch	1.46	1.59	1.66	1.49	0.13
Δ Av. Earnings (%)	-	0.24	1.01	0.08	2.41
			High SES		
HE participation	0.48	0.47	0.48	0.48	0.46
STEM share	0.42	0.42	0.35	0.41	0.42
Undermatch	1.24	1.22	1.18	1.24	1.86
Δ Av. Earnings (%)	-	-0.32	-0.90	-0.25	-3.52
SES earnings gap	34.4%	33.2%	31.1%	33.6%	26.6%

Conclusion

- We develop a model of the HE market and life cycle earnings
- Our model captures accurately empirical sorting patterns and reduced form estimates of the impacts of the 2012 reform
- We find that SES remains a strong determinant of HE sorting and earnings even accounting for the skills of students and the characteristics of programmes
- We use the model to better understand impact of 2012 reforms
- ... And to exploit alternative policy designs
- We find only muted impacts to supply side reforms
- ... But aggressive demand-side reforms can make a difference

How did the 2012 reform affect enrolment?



Lifetime student loan repayments

Cost increased the most among the highest paid

back

IV first stage

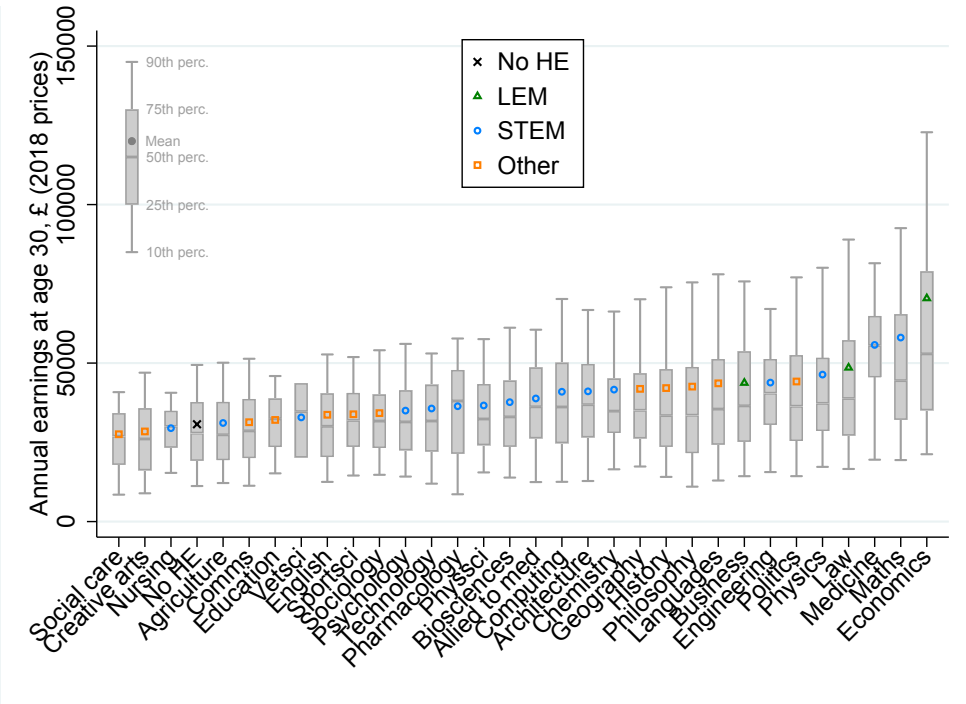
	Study STEM	Study LEM	Study Soc Sc	Quality
	Distance to closest Russell group programme			
STEM	-0.002*	0.000*	-0.003*	-0.008*
LEM	-0.003*	-0.001*	-0.003*	-0.034*
Social Sc	-0.004*	-0.002*	-0.005*	-0.015*
	Proportion of peers taking subject			
STEM	0.009*	0.002*	-0.005*	-0.033*
LEM	0.010*	0.016*	-0.001*	0.090*
F-stat	943.8	1707.5	280.3	1678.3
N		2,374,368		
X: gender, SES, skills	✓	✓	✓	✓

* significant at 1% level.

back

Earnings distribution

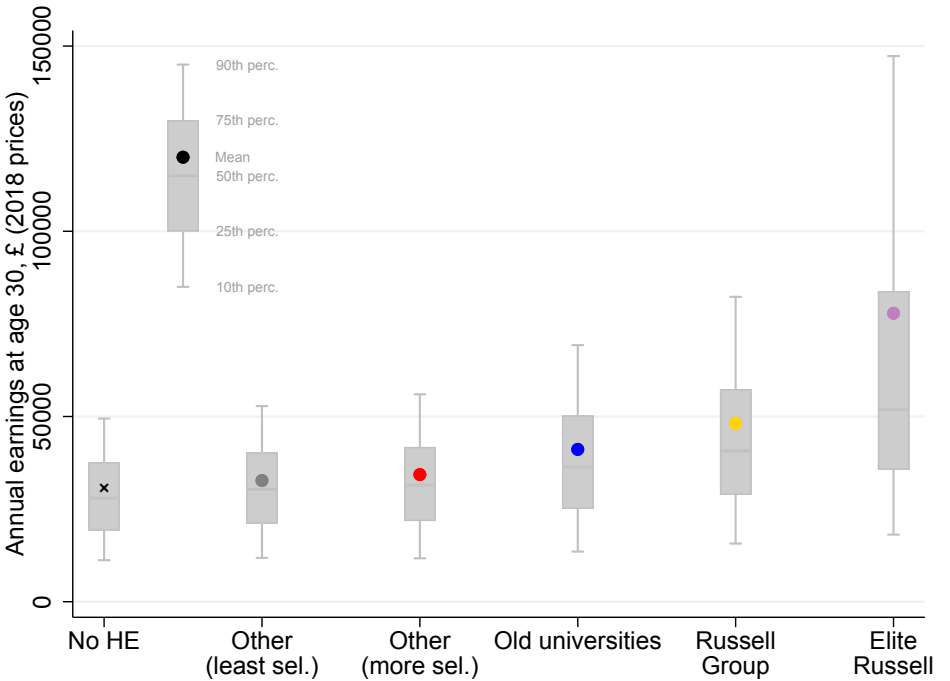
Annual earnings at 30, by subject choice



back

Earnings distribution

Annual earnings at 30, by HE institute



back