

Who goes to university and why?

A comparison across cohorts (1985 – 2015)

Oliver Cassagneau-Francis

24th April 2020

Sciences Po

Introduction

Higher education in the UK

Model and data

Identification and estimation

Results

Research questions

How important are earnings expectations for 16-18 year olds when deciding whether to go to university?

- How much do earnings expectations vary?
- How do they compare to the other factors that potential students consider? Heckman et al. (2006)'s "psychic costs"

How have these factors changed between the 1980s and today?

- Has the expected graduate-wage premium changed? And its variance?
- What about the relative importance of other factors?

What is driving the SES-gap in education attainment?

- Differences in earnings expectations?
- Other factors (e.g. prior attainment, social norms)?

Why care about the decision to attend HE?

Lots of evidence of benefits of a university degree

- Increased earnings
- Better health
- Lower levels of crime

Though often taxpayers are paying for these benefits

- £17 billion upfront costs of higher education in the UK in 2017 (Dearden et al., 2017)
- OECD countries' public spending on HE \approx 1% of GDP (OECD, 2020)

Despite growth in HE still clear inequalities in attainment

- and hence in beneficiaries of this public spending
- gap in attainment by parental education up to 30pp (OECD, 2018)

Understanding the factors that influence educational attainment is key, not only for educational outcomes but also for wider issues such as inequality and beneficiaries of public funding

- Model the (binary) decision to continue to university
- **Estimate model separately for cohorts born in 1970 and 1990**
 - Schooling, background, and subjective data from before decision
 - Earnings and occupation data after entry to labour market
 - Exploit subjective data on students' attitudes and lifestyles at 16 to better capture non-earnings factors
- **Exploit model structure to give distributions of factors in decision**
 - Compare distributions of earnings vs other factors *within cohorts*
 - Compare relative importance of factors *across cohorts*
 - Compare distributions for different demographics
- **Use subjective data to decompose “psychic costs”**
 - In progress

Importance of earnings versus other factors

- Main contributor to SES gap is non-earnings factors
 - 1970: sd of other factors twice that of earnings
 - 1990: sd of other factors more than 4 times that of earnings

Changes in factors between 1980s and today

- Increase in HE driven entirely by increase in non-earnings factors
 - earnings: \downarrow 15 p.p.; other factors: \uparrow 36 p.p.
 - mean increase in non-earnings factors \approx 50% wage increase

Contributions to SES gap

- Again non-earnings factors more important than earnings

Estimating *ex ante* returns to education

- rational expectations: e.g. Cunha and Heckman (2007b,a); Heckman et al. (2006)
- surveys to elicit expectations: e.g. Zafar (2009); Delavande and Zafar (2017); Wiswall and Zafar (2015); Arcidiacono et al. (2019); Boneva and Rauh (2019); Wiswall and Zafar (2016)

UK higher education expansion

- 'educational inequality': Blanden and Machin (2004)
- technological change: Blundell et al. (2018)
- wage premium: Walker and Zhu (2008); Green et al. (2016)

Determinants of educational choices

- Cameron and Heckman (1993)

Estimating *ex post* returns to education

- Bhuller et al. (2014)

Roy models with unobserved heterogeneity

- 'extended': D'Haultfoeuille and Maurel (2013)
- 'generalized': Abbring and Heckman (2007); Heckman et al. (2006); Heckman and Navarro (2007); Heckman and Vytlačil (2007)

Expectations (need to review)

- Mackowiak et al. (2018); Fuster et al. (2019)

Introduction

Higher education in the UK

- Growth in HE attainment

- Graduate-wage premium

- Attainment by parents' income

Model and data

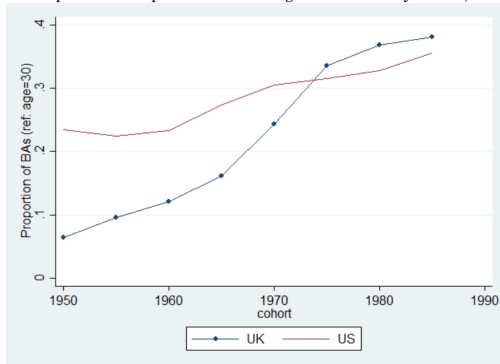
Identification and estimation

Results

Huge growth in HE attainment in the UK

- Percentage of cohort (age 30) with a degree grew from under 10% to nearly 40% in 35 years

Figure 2: Proportion of People with a BA or Higher Education by Cohort, UK and US

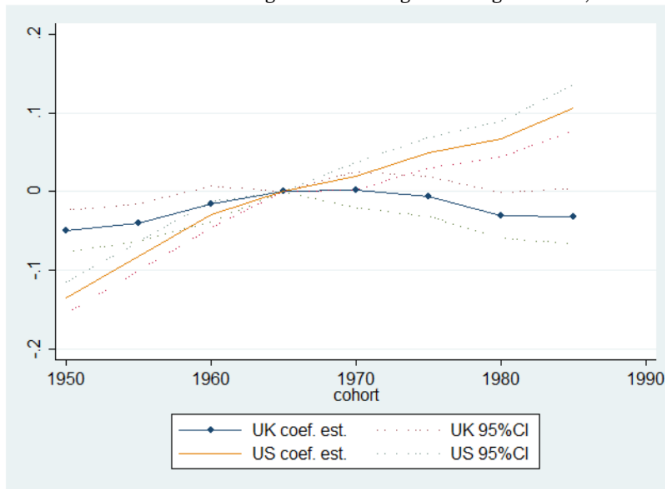


Note: sample restricted to age 22-59 and excludes full-time students. Each education-cohort cell has at least 100 observations.

Source: Blundell et al. (2018). Data from LFS (UK) and CPS (US).

Evolution of the graduate-wage premium in UK

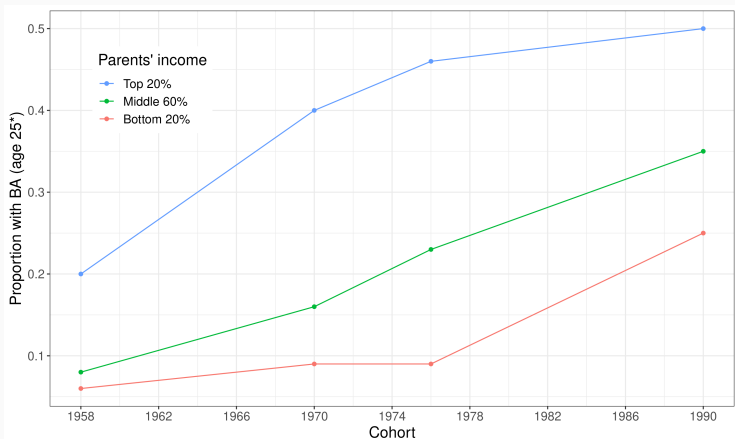
Figure 3: Ratio of BA median wage to that of high-school graduates, cohort effects



Source: Blundell et al. (2018). Data from LFS (UK) and unknown (US).

Proportion with BA by parental income

- Large and persistent SES-gap in educational attainment



Source: 1958 (NCDS) and 1976 (BHPS) from Blanden and Machin (2004). 1970 (BCS) and 1990 (LSYPE) author's own calculations. * BA proportions in 1958 and 1976 are at age 23.

Huge expansion of HE in the UK since 1970s

- Over 80% increase in proportion with a BA at 26 in 20 years
- Proportions by highest qualification at 26:
 - In 1996: O-Levels 38%, BA 19%
 - In 2015: GCSEs 22%, BA 35%

UK graduate wage premium remained constant over this period

- Ratio BA:HS median wage very similar for 1950 cohort vs 1985
- Contrast with steady growth of wage premium in the US

Despite expansion SES gap in HE attainment still high

- 50% of those in top income quintile at 16 hold BA at 25
- Only 25% of those in bottom income quintile hold a BA

Introduction

Higher education in the UK

Model and data

Identification and estimation

Results

A model of educational choice

Use a simple framework to help understand the **decision to attend university**, the generalised Roy model (Heckman et al., 2006):

$$S \equiv \mathbb{1}\{SEU_1 > SEU_0\}$$

- Individual has **subjective preferences** for different levels of schooling: university ($S = 1$) and A-levels ($S = 0$)
- Represent these by a **subjective expected utility** function (SEU)
- Choose university if $SEU_1 > SEU_0 \Rightarrow S = 1$
- And define $SEU_k \equiv s(Y_k^{ea}, \theta)$, $\forall k \in \{0, 1\}$ with
 - Y_k^{ea} (expected) *ex ante* earnings given choice k
 - θ non-earnings factors (“psychic costs”)
→ aptitude and tastes for aspects of HE

Two UK longitudinal cohort studies run by CLS at UCL

- British Cohort Study (BCS) members born in 1970
- Next Steps (LSYPE1) members born in 1989/1990

Common features:

- Follow 17,000 people from respective birth cohorts
- Surveys at age 16 and 25
 - (16) Detailed information on schooling, family background
 - (16) Subjective questions about life and future
 - (25) Earnings, occupation and qualifications
- Possibility for longer panel as both surveys ongoing

Introduction

Higher education in the UK

Model and data

Identification and estimation

Results

Additional assumptions

We need more assumptions to identify the contribution of *ex ante earnings* and *other factors* to the HE decision.

- A model for *expected ex ante earnings* given information, \mathcal{I} :

$$Y_k^{ea} \equiv \mathbb{E}[Y_k|\mathcal{I}] = X' \beta_k$$

i.e. $\mathcal{I} \subseteq X$ and $\mathbb{E}[Y_k|X]$ linear-in-parameters.

- Impose structure on the *subject expected utility* of choice k :

$$SEU_k \equiv s(Y_k^{ea}, \theta) = \alpha Y_k^{ea} + \theta' \gamma_k + \epsilon_k$$

i.e. $s(\cdot)$ linear-in-parameters.

- So that

$$S = \mathbb{1}\{\alpha(Y_1^{ea} - Y_0^{ea}) + \theta'(\gamma_1 - \gamma_0) > \epsilon_0 - \epsilon_1\}$$

A note on “earnings expectations”

The model of earnings expectations assumes rational expectations, as it is the expected value of *realised future earnings* conditional on information held by the students at the time they make their decision. This could be wrong in two important ways:

- (i) we may not observe all the information the students use to predict earnings (or alternatively we may assume they use information that they didn't have or use);
- (ii) they may make errors in their forecasting—i.e. deviate from rational expectations, despite our information set being correct

Unfortunately there is not much we can do in the way of verification or robustness checks... <add some citations that use a similar model>

Measuring θ

For most of the analysis we just want to compare the importance **earnings** and **non-earnings** factors in the decision

- i.e. $\alpha(Y_1^{ea} - Y_0^{ea})$ versus $\theta'(\gamma_1 - \gamma_0)$
- However, we do not observe θ

Therefore use *subjective questions on life and future* to proxy for θ

- Answers to questions correlated with θ that we want to measure
- Use variation in *questions* to capture variation in θ

Replace $\theta'(\gamma_1 - \gamma_0)$ with $Z'(\delta_1 - \delta_0)$ where

- Z contains questions and other characteristics
- and although $\theta \neq Z$, and $\gamma_k \neq \delta_k$, $\theta'(\gamma_1 - \gamma_0) \equiv Z'(\delta_1 - \delta_0)$

Parameter estimates ($\hat{\alpha}$, $\hat{\beta}_k$, $\hat{\Delta\delta} \equiv \widehat{\delta_1 - \delta_0}$)

- Assuming $\epsilon_0 - \epsilon_1 \sim \text{Logit}$ allows us to obtain estimates $\hat{\alpha}$, $\hat{\Delta\delta}$ by logit regression, given Y_k^{ea} and Z .
- estimate earnings as $\widehat{Y_k^{ea}} = X' \hat{\beta}_k$, from OLS of Y_k on X

Choosing X : characteristics that affect the HE decision through earnings expectations at age 16

- parents': occupations, ethnicity group, education, income
- no. A-levels taking; gender; whether think high pay is important

Introduction

Higher education in the UK

Model and data

Identification and estimation

Results

Within and between cohorts: full sample analysis

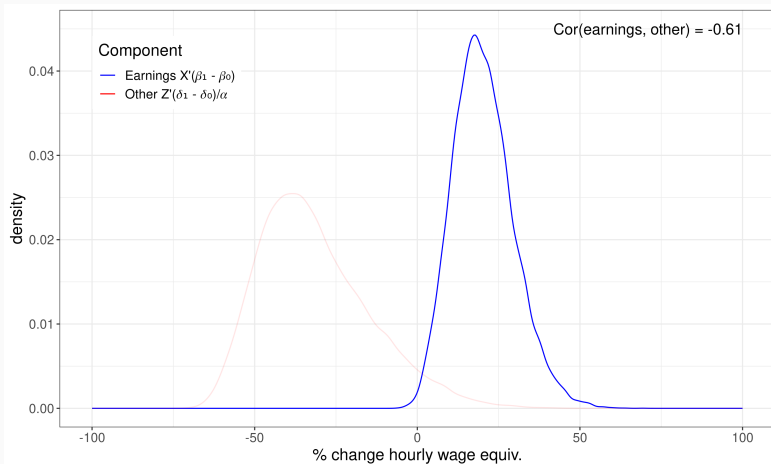
Decomposing changes in earnings across cohorts

Comparing earnings with other factors

- Compare relative importance of earnings versus other factors
- Strategy:
 - Estimate parameters $\alpha, \beta_k, \Delta\delta$
 - Use estimates and sample covariates to estimate distributions of university-wage premium, and non-earnings factors.
 - Transform these factors so they are equivalent to $\%\Delta$ in earnings
 - Plot the transformed distributions for each cohort and demographics within cohorts
- We can then compare:
 - earnings vs non-earnings factors within cohorts
 - the relative importance of these factors across cohorts

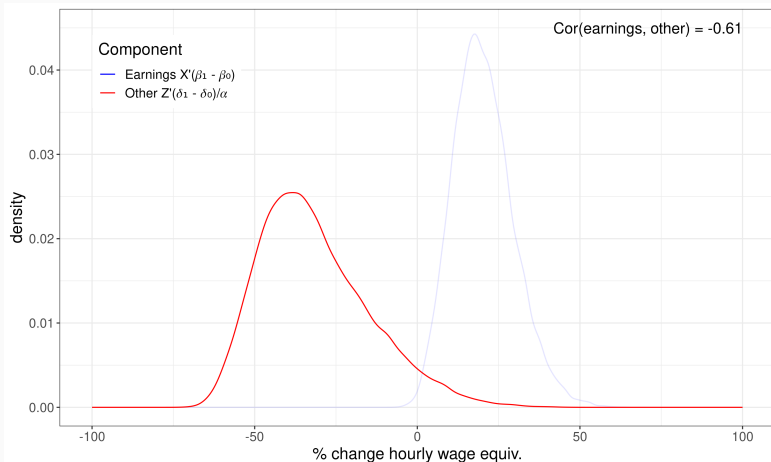
Full sample (1970)

- Earnings expectations exclusively positive (mean 20%)



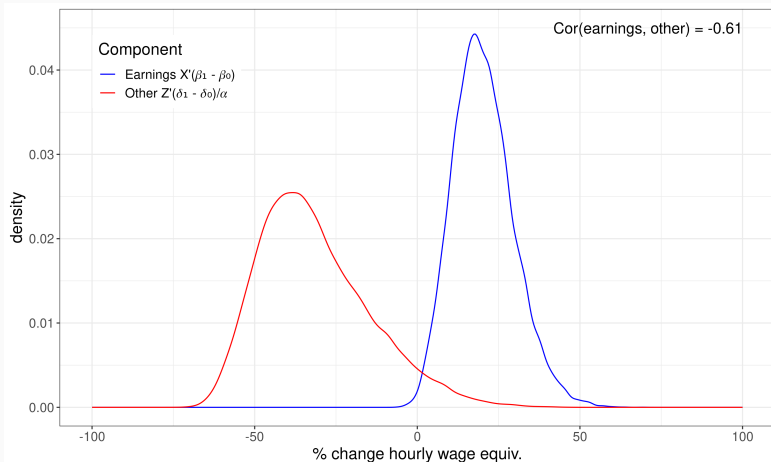
Full sample (1970)

- Earnings expectations exclusively positive (mean 20%)
- Other factors (almost) exclusively negative (mean -31%)



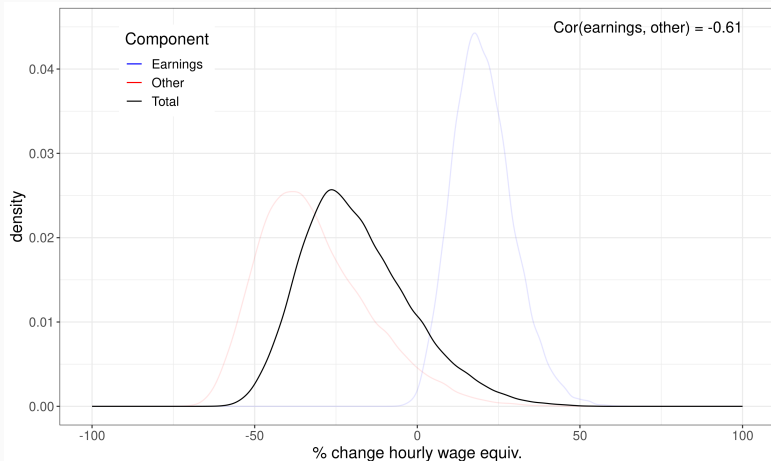
Full sample (1970)

- Earnings expectations exclusively positive (mean 20%)
- Other factors (almost) exclusively negative (mean -31%)
- $\text{sd}(\text{earnings})$ approx. half $\text{sd}(\text{other factors})$ (9.3 vs 17.0)



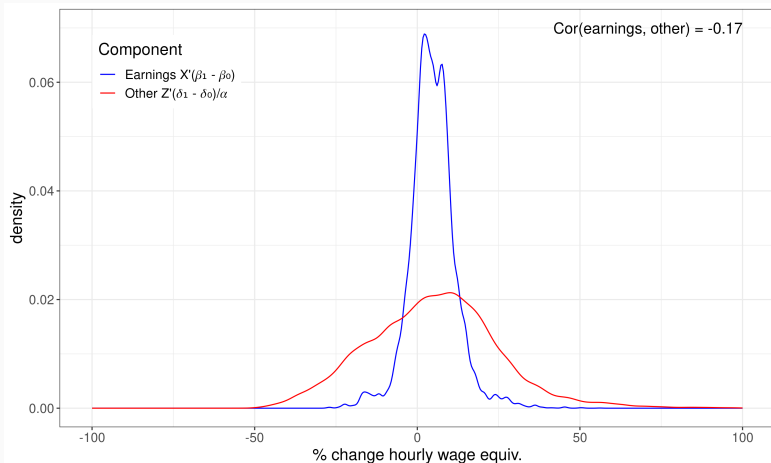
Full sample (1970)

- Earnings expectations exclusively positive (mean 20%)
- Other factors (almost) exclusively negative (mean -31%)
- $\text{sd}(\text{earnings})$ approx. half $\text{sd}(\text{other factors})$ (9.3 vs 17.0)
- Larger (abs.) mean and variance \Rightarrow other factors more important



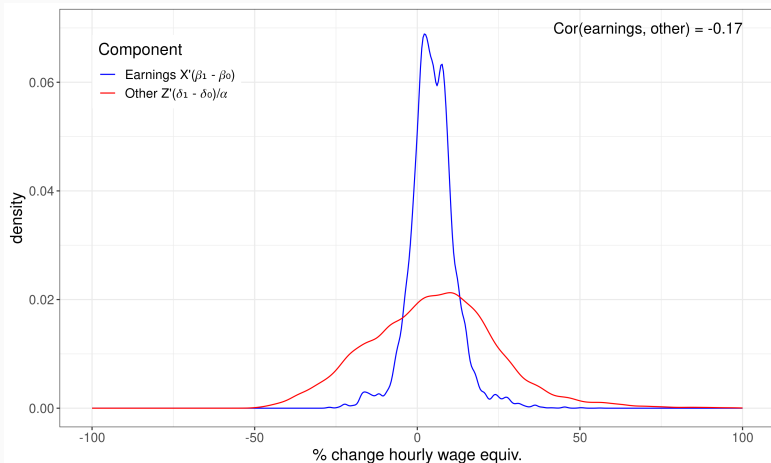
Full sample (1990)

- Both distributions similarly located, positive means (4.8% vs 5.6%)



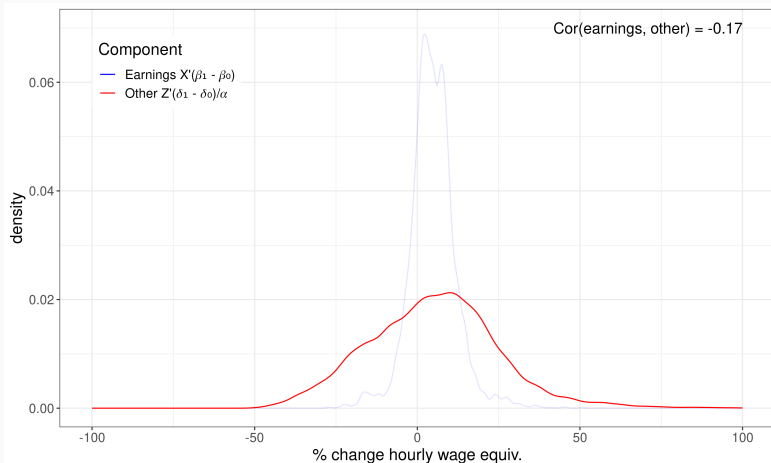
Full sample (1990)

- Both distributions similarly located, positive means (4.8% vs 5.6%)
- Earnings distribution has a much lower std dev. (7.1% vs 41%)



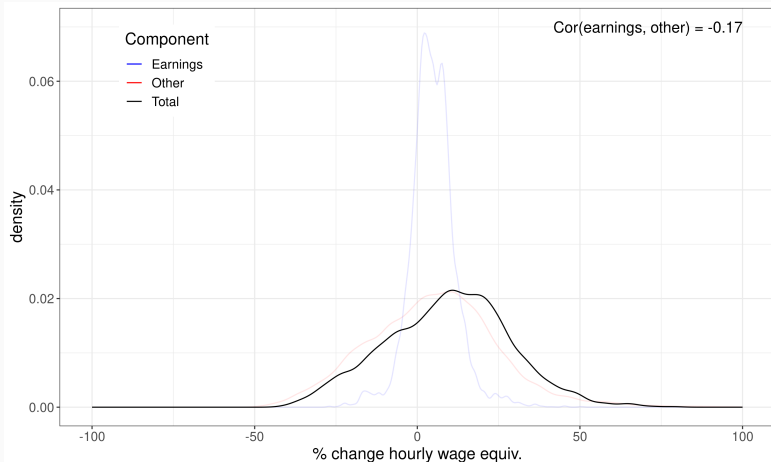
Full sample (1990)

- Both distributions similarly located, positive means (4.8% vs 5.6%)
- Earnings distribution has a much lower std dev. (7.1% vs 41%)
- Other factors distribution possesses significant *positive skew*



Full sample (1990)

- Both distributions similarly located, positive means (4.8% vs 5.6%)
- Earnings distribution has a much lower std dev. (7.1% vs 41%)
- **Other factors distribution possesses significant positive skew**
- Similar impact on rate, other factors more important for variation



Summary and discussion of full sample results

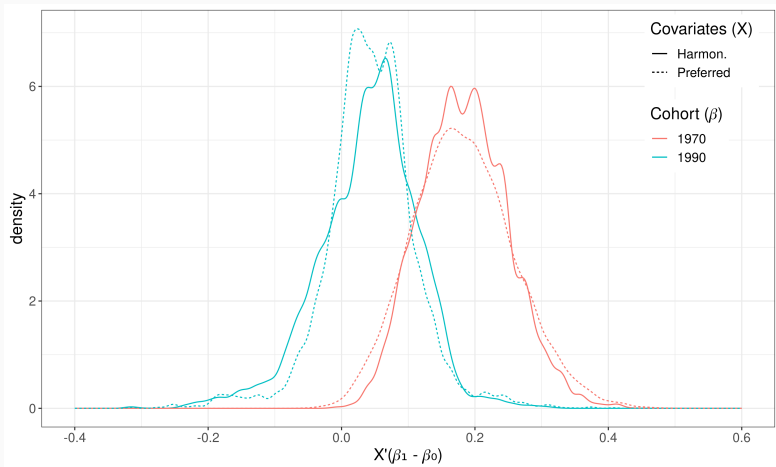
- Shape of earnings and other factors distributions quite different within cohorts:
 - earnings symmetric, low variance
 - other positive skew, larger variance
- In 1970, earnings generally positive and other factors negative
- Moving to 1990 we see:
 - a slight decrease in earnings expectations (still mostly +ve)
 - and a large increase in other factors (-ve → mostly +ve)
- In both cohorts other factors are responsible for more of the variance in HE attainment (higher variance)
- It is also the changing location of the other factors distribution that drives the large increase in attainment

Decomposing changes in earnings across cohorts

- Now focus only on earnings expectations, and estimate the same specification across both cohorts.
- Requires some combining of groups to harmonise variables across cohorts
- Strategy
 - Estimate identical specification wage model for both cohorts
 - Calculate university wage premium for each cohort
 - Calculate premium for each cohort *using estimated parameters from the other cohort*, i.e. $X_{1970}(\beta_{1,1990} - \beta_{0,1990})$
- Allows us to decompose changes across cohorts into
 - structural changes (changes in $\beta_1 - \beta_0 \equiv \Delta\beta$)
 - compositional changes (changes in distribution of X within cohort)

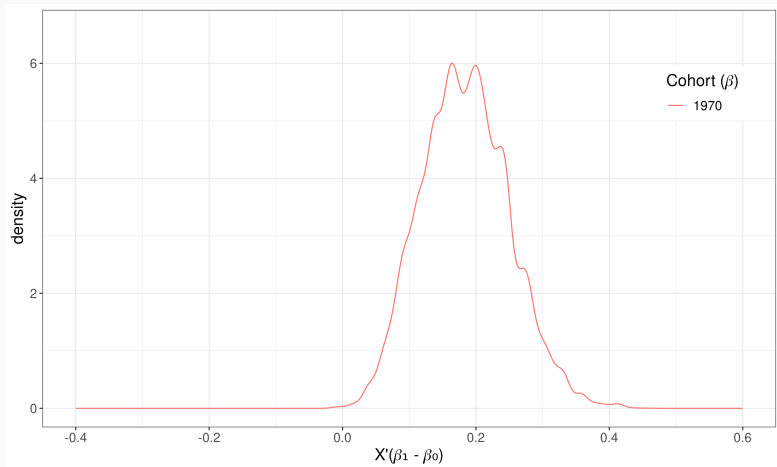
Decomposing earnings expectations: 1970 vs 1990

- “Harmonising” parameters does not affect the distribution estimates very much



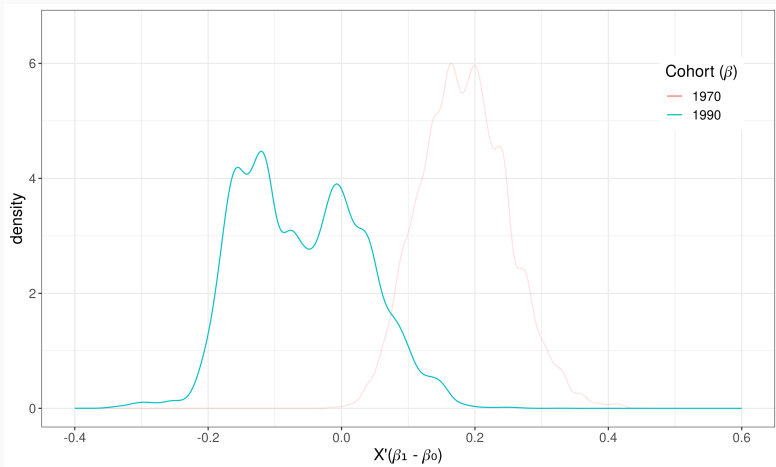
Decomposing earnings expectations: 1970 vs 1990

- Starting from β_{1970} and X_{1970}



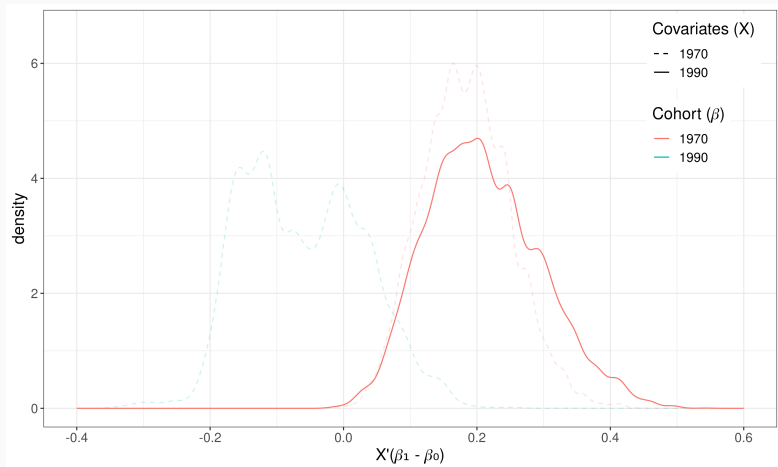
Decomposing earnings expectations: 1970 vs 1990

- Changing to β_{1990} induces a large decrease in earnings expectations...



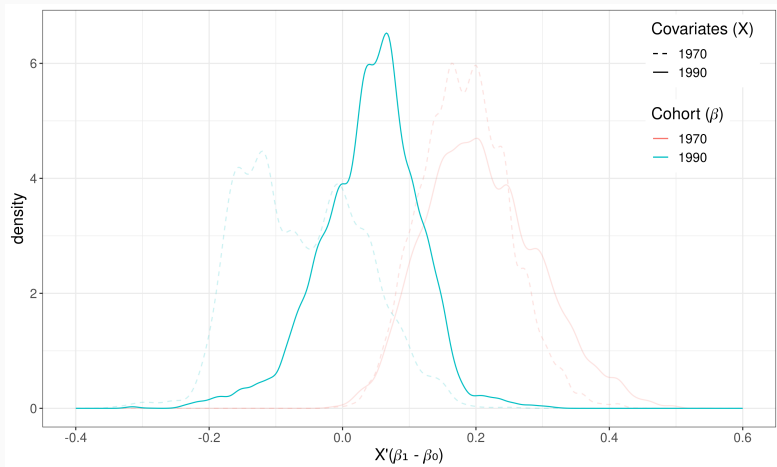
Decomposing earnings expectations: 1970 vs 1990

- ...while changing to X_{1990} (β_{1970}) has little effect.



Decomposing earnings expectations: 1970 vs 1990

- Changing both X and β results in reduced expectations, though by less than just changing β .



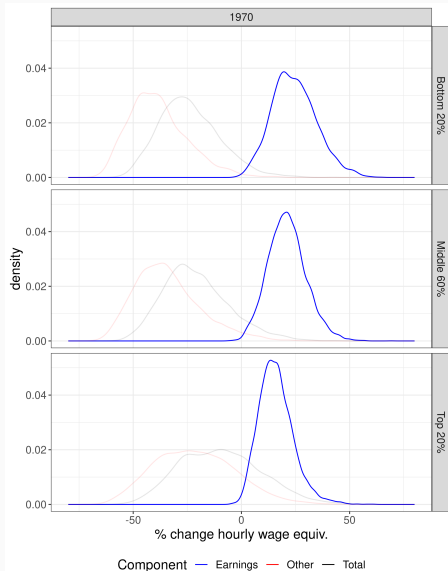
Summary of decomposition

- Purpose of exercise is to compare the effect of changes in parameters versus changes in X on earnings expectations
- This allows two “counterfactual scenarios”:
 - moving the 1970 cohort to 1990 ($X'_{1970}\Delta\beta_{1990}$)
 - moving the 1990 cohort to 1970 ($X'_{1990}\Delta\beta_{1970}$)
- We see that β_{1990} result in lower expectations than β_{1970}
- Changing $\beta_{1970} \rightarrow \beta_{1990}$ results in large decrease in expectations (X_{1970}), while $\beta_{1990} \rightarrow \beta_{1970}$ (X_{1990}) results in increase
- Change in composition ($X_{1970} \rightarrow X_{1990}$) tempers the change due to parameters so the final distribution lies between $X'_{1970}\Delta\beta_{1970}$ and $X'_{1970}\Delta\beta_{1990}$

Comparing factors across SES groups

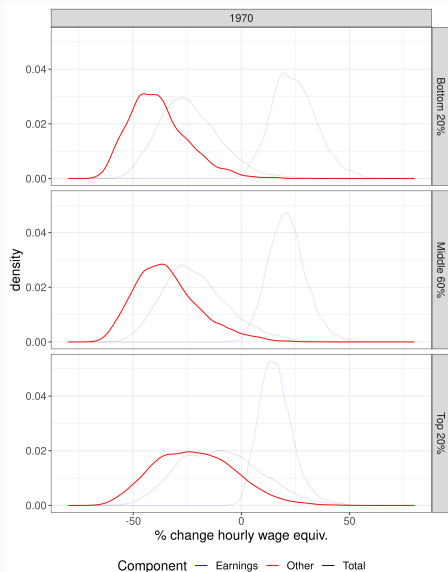
- Aim of exercise is to compare factors in the HE decision across SES groups, and uncover the cause of the SES gap
- Given the longitudinal nature of the surveys, we have detailed information on the family background of students at 16
- Lack a consistent definition of SES, so use parental income
- The following slides present the distributions of earnings expectations and other factors, *conditional on parental income quintiles at 16*
- The parameters are constant across the cohort, but each income group contains only the X, Z of individuals in that group
- **Extension:** re-estimate the model on subsamples by parental income

By parents' income (1970)



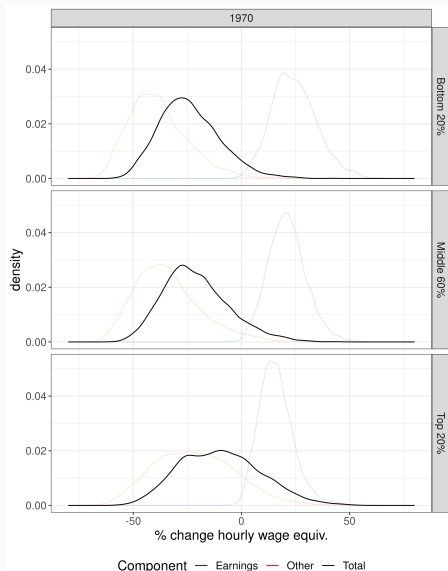
- Mean earnings expectations decreasing in parental income, variance also decreasing

By parents' income (1970)



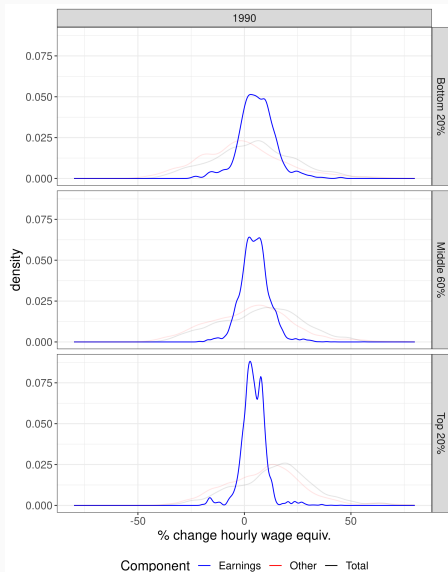
- Mean earnings expectations decreasing in parental income, variance also decreasing
- Other factors very similar distribution for low- and middle-income, higher mean and variance for top 20%

By parents' income (1970)



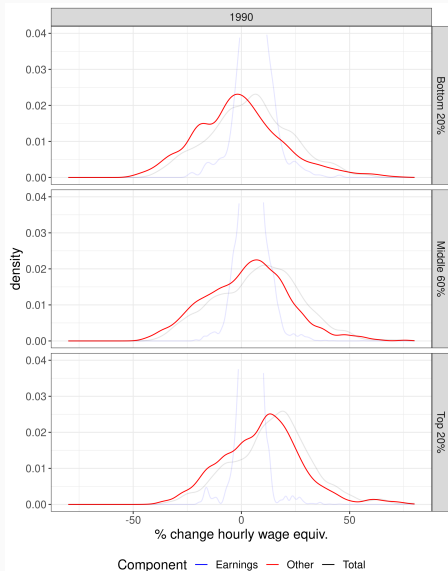
- Mean earnings expectations decreasing in parental income, variance also decreasing
- Other factors very similar distribution for low- and middle-income, higher mean and variance for top 20%
- Combined shows the importance of other factors, and can clearly see HE rate increasing with income

By parents' income (1990)



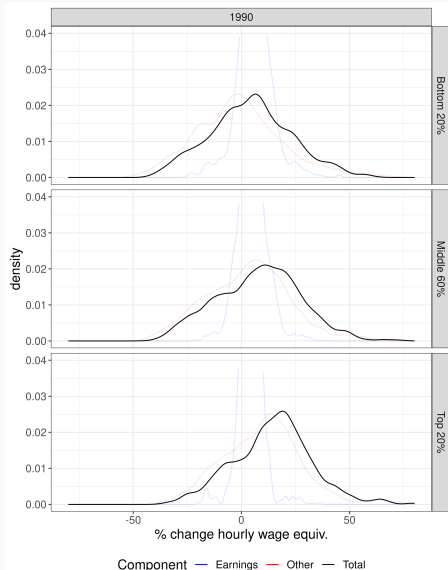
- Mean earnings expectations similar across income groups, variance decreasing in income

By parents' income (1990)



- Mean earnings expectations similar across income groups, variance decreasing in income
- Other factors have very similar-shaped distributions across groups, mean increasing in income

By parents' income (1990)



- Mean earnings expectations similar across income groups, variance decreasing in income
- Other factors have very similar-shaped distributions across groups, mean increasing in income
- Again combining earnings and other factors shows the importance of other factors

Summary of SES analysis

- In both cohorts, other factors are the main contributor to the SES gap in HE attainment
- Mean earnings expectations actually slightly decreasing in income for 1970 cohort
- Though the variance of earnings expectations is decreasing in income for both cohorts
- Meanwhile, clear that other factors contribution to the HE decision is increasing in earnings
- Further work:
 - Robustness check estimating model separately on subsamples (allow decomposition analysis)
 - Understand the main components of other factors (“psychic costs”)

- Decompose other factors into meaningful components of “psychic costs”
 - Create a mapping, $f : Z \rightarrow \theta$ (via PCA or manual classification)
 - Re-estimate model and redo analysis on components of θ
- Allow for unobserved heterogeneity in earnings
 - May be missing information students use to forecast their earnings
 - Following Heckman et al. (2006), allow for this missing data by explicitly including it in the model of earnings expectations

- Summary of findings so far
 - Earnings expectations (as measured here) are much less important than other factors in the decision to continue to higher education (in line with recent results from other research)
 - Over a period of huge growth in HE attainment, earnings expectations in the UK have fallen slightly and it is other factors that have driven this increase
 - A similar story explains the continuing SES gap in attainment: earnings expectations are broadly similar across parental income groups, and other factors are the culprit
- Next steps
 - Decompose other factors into meaningful components

References i

- Abbring, J. H. and Heckman, J. J. (2007). Chapter 72 Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. In *Handbook of Econometrics*, volume 6, pages 5145–5303. Elsevier.
- Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2019). Ex Ante Returns and Occupational Choice. page 47.
- Bhuller, M., Mogstad, M., and Salvanes, K. G. (2014). Life Cycle Earnings, Education Premiums and Internal Rates of Return. page 34.
- Blanden, J. and Machin, S. (2004). Educational Inequality and the Expansion of UK Higher Education. *Scottish Journal of Political Economy*, 51(2):230–249.
- Blundell, R., Green, D., and Jin, W. (2018). The UK Education Expansion and Technological Change. page 71.
- Boneva, T. and Rauh, C. (2019). Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns. *JPE*, page 83.
- Cameron, S. V. and Heckman, J. J. (1993). Determinants of Young Male Schooling and Training Choices. Working Paper 4327, National Bureau of Economic Research.
- Cunha, F. and Heckman, J. (2007a). The Evolution of Labor Earnings Risk in the U.S. Economy. page 53.

- Cunha, F. and Heckman, J. J. (2007b). Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling. *Labour Economics*, 14(6):870–893.
- Dearden, L., van der Erve, L., Belfield, C., and Britton, J. (2017). Higher Education funding in England: past, present and options for the future. Technical report. ISBN: 9781911102489.
- Delavande, A. and Zafar, B. (2017). University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints. page 56.
- D'Haultfoeulle, X. and Maurel, A. (2013). Inference on an extended Roy model, with an application to schooling decisions in France. *Journal of Econometrics*, 174(2):95–106.
- Fuster, A., Perez-Truglia, R., and Zafar, B. (2019). Expectations with Endogenous Information Acquisition: An Experimental Investigation. page 68.
- Green, D. A., Jin, W. M., and Blundell, R. (2016). The UK wage premium puzzle: how did a large increase in university graduates leave the education premium unchanged? Working Paper Series, IFS. Series: Working Paper Series.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In *Handbook of the Economics of Education*, volume 1, pages 307–458. Elsevier.

- Heckman, J. J. and Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136(2):341–396.
- Heckman, J. J. and Vytlacil, E. J. (2007). Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6, pages 4875–5143. Elsevier.
- Mackowiak, B., Matejka, F., and Wiederholt, M. (2018). Survey: Rational Inattention, a Disciplined Behavioral Model. page 42.
- OECD (2018). *Equity in Education: Breaking Down Barriers to Social Mobility*. PISA. OECD.
- OECD (2020). Public spending on education (indicator). *Education at a glance: Educational finance indicators*.
- Walker, I. and Zhu, Y. (2008). The College Wage Premium and the Expansion of Higher Education in the UK. *The Scandinavian Journal of Economics*, 110(4):695–709.
- Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2016). Human Capital Investments and Expectations about Career and Family. Technical Report w22543, National Bureau of Economic Research, Cambridge, MA.

Zafar, B. (2009). How do College Students Form Expectations? *SSRN Electronic Journal*.