The Role of Earnings, Financial, and other Factors in University Attendance

Oliver Cassagneau-Francis

The role of earnings, financial, and other factors in university attendance

Oliver Cassagneau-Francis*

24th May 2021
(click here for the latest version)

Abstract

Why do some people choose to attend university, and enjoy state-subsidised benefits, while others do not? We shed new light on this key issue by comparing and quantifying the roles of earnings, financial, and non-pecuniary factors in the educational decisions of young people in the UK. We investigate changes in these factors over time, and their implications for social mobility. We specify a model of educational choice, explicitly including expectations about earnings, financial, and non-pecuniary factors. Our estimation strategy exploits panel survey data on young people's expectations about key outcomes both at, and after, university, linked to their realised outcomes. Income maximisation, despite its prevalent role in the literature, is only a small part of the story: other factors are four times as important as earnings in determining whether someone goes to university. Non-pecuniary factors also drive both the SES-gap in educational attainment, and the huge growth in degree attainment between the 1980s and 2010s.

Keywords: Higher education; Earnings; Psychic costs; Wage premium; Educational choice.

JEL codes: E24; I23; I26; J24

*Sciences Po, Paris; email: oliver.cassagneaufrancis@sciencespo.fr. I am grateful to my advisors Ghazala Azmat and Jean-Marc Robin for their invaluable guidance and support throughout this project. I thank Zsofia Barany, Johannes Boehm, Jeanne Commault, Pierre-Philippe Combes, Nicolo Dalvit, Michele Fioretti, Eric French, George-Levi Gayle, Kerstin Holzheu, Guy Laroque, Hugo Lhuillier, Michela Tincani, and numerous seminar participants at Sciences Po for helpful comments and discussions. I thank the Centre for Longitudinal Studies at UCL for administering the survey and, along with the UK Data Service, providing access to the data. All errors are mine.

1 Introduction

Deciding whether or not to attend university is one of the most important choices in a young person's life. Graduates enjoy higher wages, better health, and commit less crime than their less-educated peers (Heckman, Humphries, and Veramendi, 2018; Oreopoulos and Petronijevic, 2013). These benefits are often heavily subsidised, with governments in developed countries spending around 1% of their countries' GDP on higher education (OECD, 2018). Therefore, understanding young people's educational decisions is key, not only to better understand the direct effects of educational policies, but also for wider issues such as inequality and to identify the beneficiaries of public spending. Traditionally, economists have focused on higher wages, pecuniary costs, and financial frictions to explain this decision—a narrative involving comparative advantage and credit constraints. However, more recent work suggests this is not the whole story, with income maximisation proving incapable of fully explaining patterns in the data.

We study the role of both pecuniary and non-pecuniary factors in the decision to attend university, deepening our understanding of young people's educational decisions, and the factors they value. A deeper understanding of these factors helps policymakers better target policies designed to encourage young people to attend university, making them more efficient and cost-effective. Our analysis addresses two key questions concerning higher education: (i) How has the importance of these factors in the decision changed between the 1980s and today? Understanding the role of individuals' decisions in this educational expansion is important both from a historical perspective, and to inform future policy decisions. (ii) What is driving the educational attainment gap between advantaged and less-advantaged potential students? Given the apparent benefits afforded by a university degree, understanding why graduates are disproportionately from advantaged backgrounds is vital to understand the role of higher education in inequality and social mobility.

The data come from two longitudinal studies in the UK, which follow representative samples from cohorts born 20 years apart. The UK higher education system possesses a number of features which make it particularly suited to address the above questions. There is a comprehensive system of government-funded loans available to all students, with very generous income-contingent repayment conditions (Crawford and Jin, 2014), allowing us to abstract from credit constraints in our analysis. Higher education in the UK has undergone significant growth in recent decades: only 12% of the population held a degree in 1993; this had grown to over 35% by 2015. This sharp increase in university attainment was accompanied by a surprisingly flat graduate-wage premium, which hardly changed over this period (Blundell, Green, and Jin, 2021). This apparent "puzzle" highlights the (growing) importance of non-pecuniary (or certainly non-wage)

factors in educational decisions. Despite the growth in higher education, there is still a large socio-economic gap in education attainment, a gap reflected in other developed countries (OECD, 2018).

To guide our empirical analysis, we specify a parsimonious model of educational choice in the spirit of Roy (1951) explicitly including both earnings and other (chiefly non-pecuniary) factors. Models of this type have been applied in recent years to educational choice, a literature begun by a series of papers by Flavio Cunha, James Heckman and coauthors (Cunha, Heckman, and Navarro, 2004; Cunha and Heckman, 2007; Heckman, Lochner, and Todd, 2006). These and more recent papers highlight the importance of non-pecuniary factors (often called "psychic costs") in explaining educational choices, both at the intensive (e.g. major choices in Wiswall and Zafar (2015)) and extensive margins (D'Haultfoeuille and Maurel, 2013; Boneva and Rauh, 2020). Our contributions to this literature are threefold. First, we compare and quantify earnings and other factors[1] in the decision to attend university, both across socio-economic groups and over time. As far as we are aware, we are the first to analyse the role of non-pecuniary factors over time in educational decisions. Second, we exploit information on realised earnings and choices, and elicited expectations about other factors. A growing literature studies young people's choices by eliciting expecations about future earnings from students, but not about non-pecuniary factors.[2] For much of this important work realised outcomes are not (yet) available.[3] And third, we employ longitudinal data from two large and representative samples from cohorts born 20 years apart. Prior work using elicited expectations have often used smaller, selected samples, either from a single US college (Arcidiacono et al., 2020) or self-selected survey respondents (Boneva and Rauh, 2020).

Our empirical strategy faces two chief difficulties: (i) the model requires *expected* earnings and we observe *realised* earnings; (ii) we only observe earnings at a single point in time. To solve the first issue we follow an approach pioneered by Cunha and Heckman (2007). They show how to map realised earnings into expected earnings via assumptions of rational expectations and on the contents of agents' information sets. To solve the second issue, we assume that earnings at 25 are a sufficient statistic for the earnings students' consider when deciding whether to continue to university. After solving these issues, the model is straightforward to estimate using standard techniques from the discrete-choice literature.

We first estimate the model on longitudinal survey data from a cohort born in 1990. We combine estimated preferences with observed and estimated expectations to estimate the distributions of earnings and other factors in the decision to attend university. Although

[1] We will often group financial and non-pecuniary factors together and refer to them as "other factors".

[2] Boneva and Rauh (2020) is a notable exception.

[3] Outcomes are beginning to become available for some early work which elicited expectations, see for example Arcidiacono, Hotz, Maurel, and Romano (2020) and Gong, Stinebrickner, and Stinebrickner (2019).

the distributions of earnings and of other factors share similar shapes and locations—bell-shaped, with slightly positive means—the variance of the other factors distribution is four times higher. The similar shapes and locations of the distributions near zero mean that the chief determinant of whether or not someone decides to go to university is chiefly determined by their expectations about other factors. To show this more clearly, we study the effects of changing the values of these factors. Assigning everyone in the sample expectations about other factors equal to the 25th-percentile results in 24% attending university; assigning them values equal to the 75th-percentile results in over 99% attending. Repeating the same exercise with expectations about earnings results in 60% (25-th percentile) and 74% (75-th percentile) of people attending, a much smaller effect.

Next, we re-estimate the model on data from an earlier cohort born in 1970. Comparing the distributions of earnings and other factors from the earlier with the later cohort allows us to assess their role the higher education growth seen over this period. The distribution of the expected graduate-wage premium remained quite stable over this period, its mean and variance decreasing slightly. Other factors, however, changed drastically, shifting right so that the strongly negative mean of the 1970 cohort becomes slightly positive for the 1990 cohort. The variance of other factors increased slightly too. Taken together these results suggest the increase in degree attainment—which went from 14% in the 1970 sample, to 69% in the 1990—was entirely driven by changes in non-earnings factors.

Returning to the 1990 cohort, we split the sample into three groups by socio-economic status (SES) measured by parental earnings at sixteen,[4] to investigate the role of earnings and other factors in the socio-economic gap in university attainment. We recalculate the distributions of earnings and other factors for each of the three SES groups. The distribution of expectations about the graduate premium is remarkably stable across the three groups, with means ranging from 5.9% (low SES) to 4.3% (high SES), and variances also decreasing slightly with parental income. For other factors, there is much more variation across SES: the low-SES mean is -0.7%, while the high-SES mean is 10.2%. The socio-economic gap in university attainment is entirely driven by factors other than earnings.

The rest of the paper is organised as follows. The next section discusses the context of higher education in England and presents some facts about educational attainment and the wage premium in England over recent decades. Section 3 details the two surveys that are the sources of data for estimating the model, which section 4 introduces. Section 5 discusses our empirical strategy. The results follow in section 6. In section 7, we describe the improved model, its identification and our estimation strategy. Section 8 concludes.

[4]These correspond to the bottom quintile of parental earnings in the sample (low SES), the middle-three quintiles (middle SES), and the top quintile (high SES).

2 Background and context

Higher education in England has seen substantial changes in recent decades, undergoing substantial growth and an overhaul of its funding system.[5] Figure 1 shows trends in higher education attainment and wages in the UK, highlighting key differences with the US. Growth in attainment has been much steeper in the UK than in the US. The proportion of UK (US) BAs in a given cohort at age 30 increased from less than 10% (25%) for those born in 1950, to nearly 40% (35%) for those born in 1985 (Blundell et al., 2021, see figure 1a). Alongside this rapid growth in attainment, the graduate wage premium has remained flat in the UK, while it has been steadily increasing in the US (Blundell et al., 2021, see figure 1b). Finally, the introduction of tuition fees meant university went from being completely free prior to 1998, to costing over £3,000 a year in 2006.[6] Such rapid growth in higher education, over a period of increased fees and stagnant returns, raises questions about what drove so many more people to attend university—questions we shed new light upon in this paper.

In figure 2 we recreate the UK trends using different survey data, and decompose them by quintiles of parental earnings at age sixteen. The top (blue) line represents those children whose parents earn in the top twenty percent of parents in their cohort and sample, the middle line capture those with parents who earn in the middle three quintiles, and the bottom line represents children with parents earning in the bottom twenty percent. There are two key features to highlight in this plot. The first is the constant growth in BA attainment for all three groups, although the "top" group grew faster at the beginning of the period and the "bottom" group saw the biggest increases since 1976. Second, despite the growth across all groups, the attainment gap between groups grew for all combinations of groups. For example, the attainment gap in the 1958 cohort was just under 15 percentage points between the top and bottom groups, but had grown to 25 percentage points for the 1990 cohort. Therefore, despite huge growth in educational attainment across socio-economic groups, it is still young people from more-advantaged backgrounds who disproportionately gain a university education.

3 UK cohort panel data

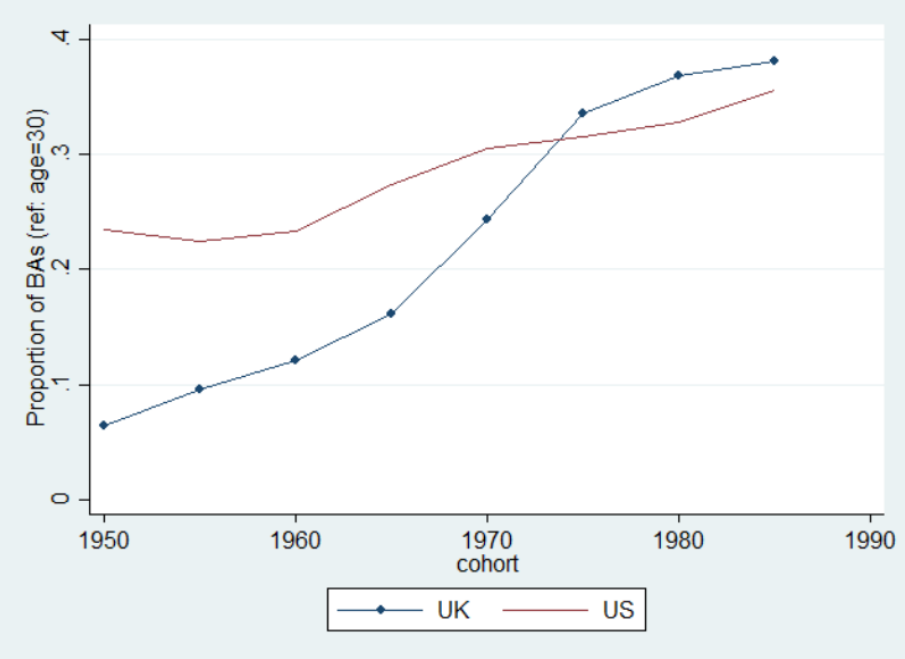
The data used in this paper come from two cohort studies following the lives of people in the UK. The primary focus is on the more recent study, Next Steps, which follows a representative sample of 15,770 people born in England in 1989 or 1990. Data collection involved annual face-to-face interviews between 2004 and 2010 (waves 1–7), plus a further

[5] These changes have been widely documented (Blanden and Machin, 2004; Walker and Zhu, 2008; Devereux and Fan, 2011; Blundell et al., 2021).

[6] A comprehensive system of loans (with repayments contingent on future income), and means-tested grants was also introduced. These are discussed in more detail in appendix A.

Figure 1: Higher education and wages in the UK vs the US in recent decades

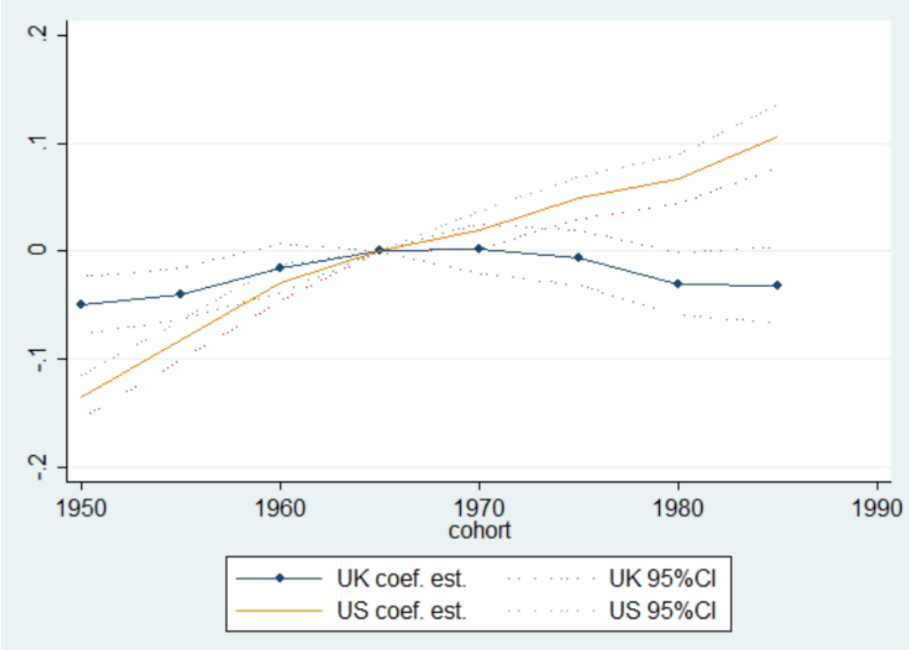
(a) Proportion of people with a BA or higher education by cohort, UK and US



Source: Blundell, Green, and Jin (2018)

Notes: Sample restricted to ages 22–59 and excludes full-time students. Each education-cohort cell has at least 100 observations.

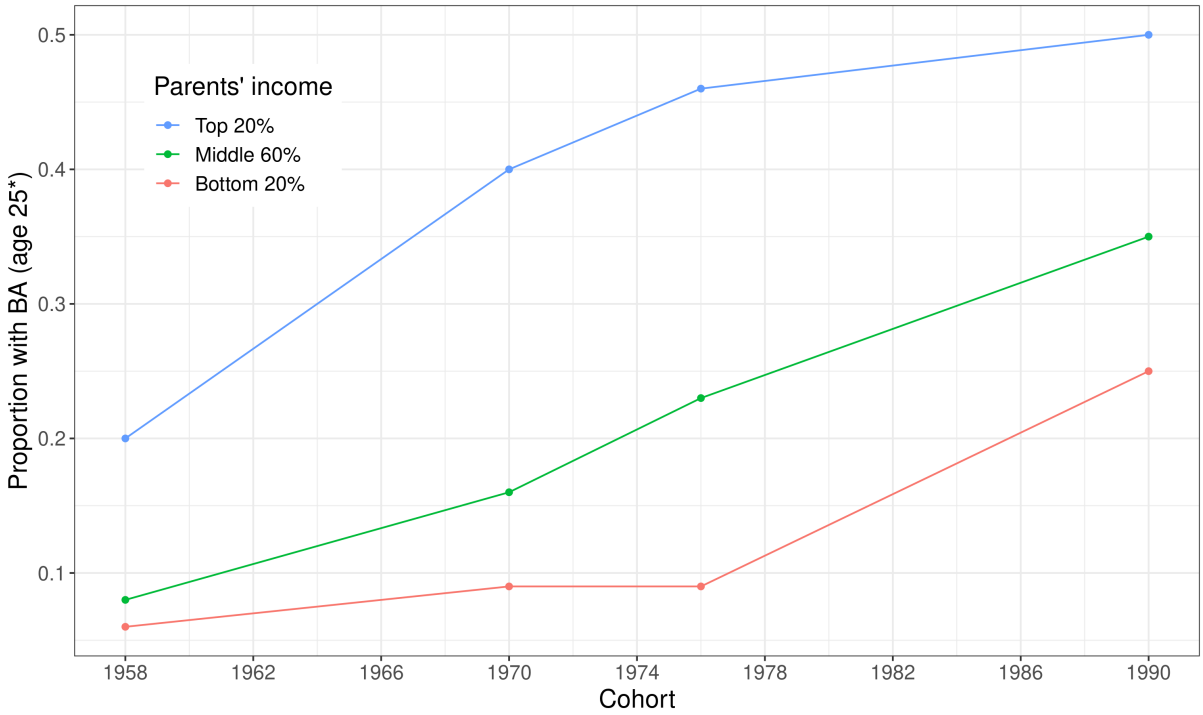
(b) Ratio of BA median wage to that of high-school graduates, cohort effects



Source: Blundell et al. (2018)

Notes: Data from LFS (UK) and unknown (US).

Figure 2: Proportion of cohort with a BA by parental income at 16



Source: 1958 (NCDS) and 1976 (BHPS) from Blanden and Machin (2004). 1970 (BCS) and 1990 (LSYPE1) author's calculations.
Notes: *BA proportions in 1958 and 1976 are at age 23.

round of interviews in 2015 (wave 8).[7] Of particular interest for the current analysis is information on: schooling, family background, and expectations about university and the future at age sixteen (before applying to university); and on earnings, occupation and qualifications at age twenty-five (after entry to the labour market). The elicited expectations provide a direct measure of students' beliefs about the future. We supplement the data from Next Steps with data from the earlier British Cohort Study (BCS) to analyse changes in factors across time. The BCS is a similar study to Next Steps, following nearly 17,000 people born in the UK in April 1970.

3.1 Descriptive statistics for Next Steps

Individual characteristics and sample selection. Table 1 presents summary statistics from the two waves of Next Steps (4 and 8) on which the analysis focuses, comparing the full sample (column 1) to those included in the analysis (column 2). Only those with a minimum of 5 GCSEs at A*-C or equivalent were asked about their expectations, information vital to the analysis in this paper.[8] The young people not asked about their

[7] The study is ongoing and the cohort members will be interviewed again in 2021, with plans to make the data available by 2023.

[8] These are referred to as "high-achieving" students in the survey documentation. Blundell et al. (2021) consider grade C at GCSE as the UK equivalent to high-school graduation in the US.

expectations are not included in the analysis. The proportion who are female increases among those who remain in the subsample, along with parental income and number of A-levels taken at age seventeen, and the mean wage and proportion who have attended university at age twenty-five. Although this reduces the sample by almost a third (first row, table 1), those omitted are likely students who would have found it very difficult to attend university. They are an important group to study, but their omission is not fatal to the current analysis.

Elicited expectations about university. Figure 3 shows the proportion of young people who mentioned certain advantages and disadvantages of attending university. Focusing first on the reported advantages in figure 3a, access to "better opportunities" and to "better jobs" were the two most common advantages of a university degree mentioned by respondents. In close third was "more qualifications", with getting a "well-paid job" in fourth place. An enjoyable "social life" rounds out the top five most popular advantages, with "learning", "personal development", and "gain life skills" also popular responses. Although some of the responses are arguably linked to higher pay, there are many that are not, for example "social life" and "personal development". In addition, the presence of "well-paid job" as a specific response suggests other career-related responses are capturing broader notions than pay alone. Turning to the disadvantages in figure 3b, the three responses mentioned most often are all financial concerns: "get into debt", "costs (general)", and "too expensive". However, many of the disadvantages mentioned reflect non-pecuniary aspects of a person's career ("no job guarantee"), or life at university ("heavy workload", "leave home"). Together these responses provide direct measures of the pecuniary, financial, and non-pecuniary factors in the decision to attend university.

3.2 The decision to attend university

As a first step towards understanding and quantifying the factors that potential students consider when deciding whether to attend university, we analyse the correlations between different possible factors and university attendance. This provides evidence on the predictive power of different factors in the decision, and hence informs which are important to include in the model. We proceed by estimating logit models of university attendance, measured by degree attainment at age twenty-five. Estimates of key parameters are in appendix B, figure B10. As these are qualitative survey responses, they are coded as indicator variables and are relative to a reference category. For the advantages and disadvantages in figure B10a the reference category is those who did not mention the corresponding advantage or disadvantage when surveyed. For the attitudes in figure B10b, the reference category is to "strongly agree" with the corresponding statement. Also included are a range of background characteristics for which we do not report the parameter

Table 1: Descriptive statistics for the full sample and the analysis subsample

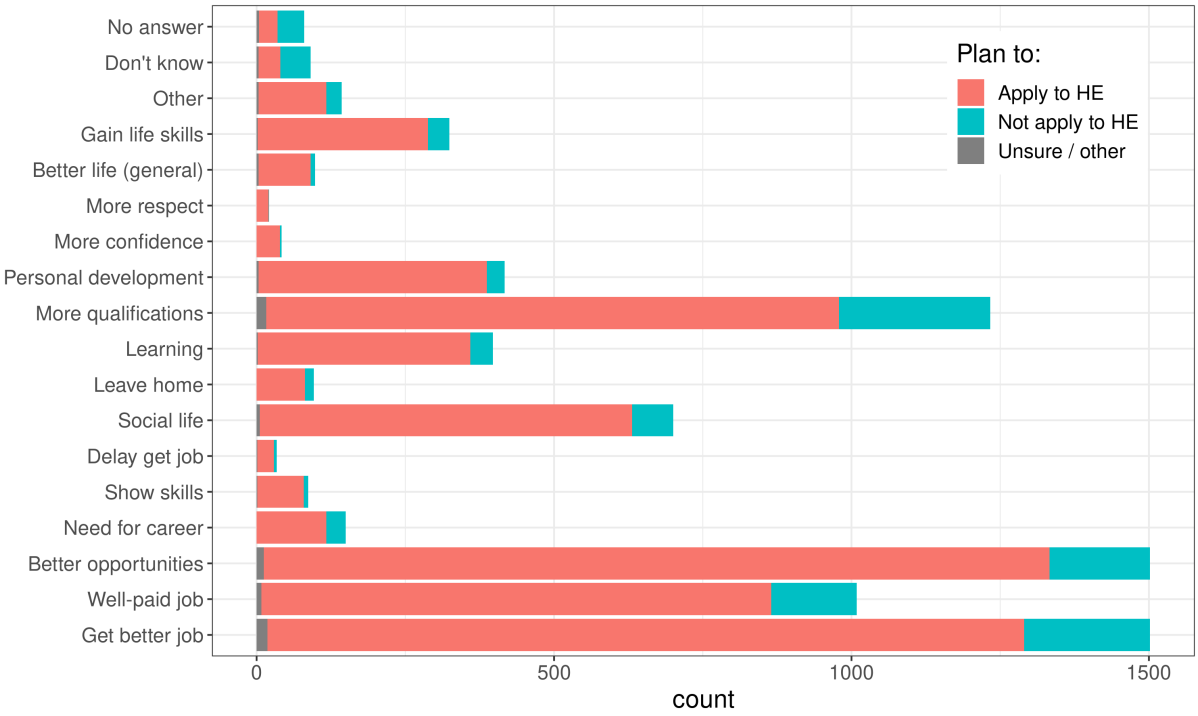
	Full sample	Subsample
N	6,628	4,640
Female	0.55	0.57
Ethnic group		
White	0.70	0.70
South Asian	0.17	0.17
Black Carib./African	0.08	0.06
Other	0.07	0.08
Main parent's occupation		
SOC 1–3	0.29	0.35
SOC 4–5	0.15	0.17
SOC 6–9	0.26	0.33
NA	0.29	0.25
Wave 4 — age sixteen		
Parental income (annual, GBP)		
20th percentile	12,500	17,500
80th percentile	48,000	58,000
Taking A-levels (age 17)	0.59	0.79
Mean #A-levels	3.71	3.79
Wave 8 — age twenty-five		
Degree	0.58	0.68
Russell group[†]	0.26	0.28
Low SES[§]	0.46	0.60
Middle SES[§]	0.56	0.65
High SES[§]	0.71	0.75
Employed	0.83	0.87
Wage (weekly, GBP)[‡]		
Median	393	424
No degree	369	403
Degree	450	461
Mean	413	445
(standard deviation)	(199)	(200)
No degree	392	424
(std dev.)	(196)	(201)
Degree	465	476
(std dev.)	(202)	(201)

[†]Among degree holders. [‡]Median wage. [§]Measured by parental income quintiles: bottom 20% (low SES); middle 60% (middle SES); and top 20% (top SES).

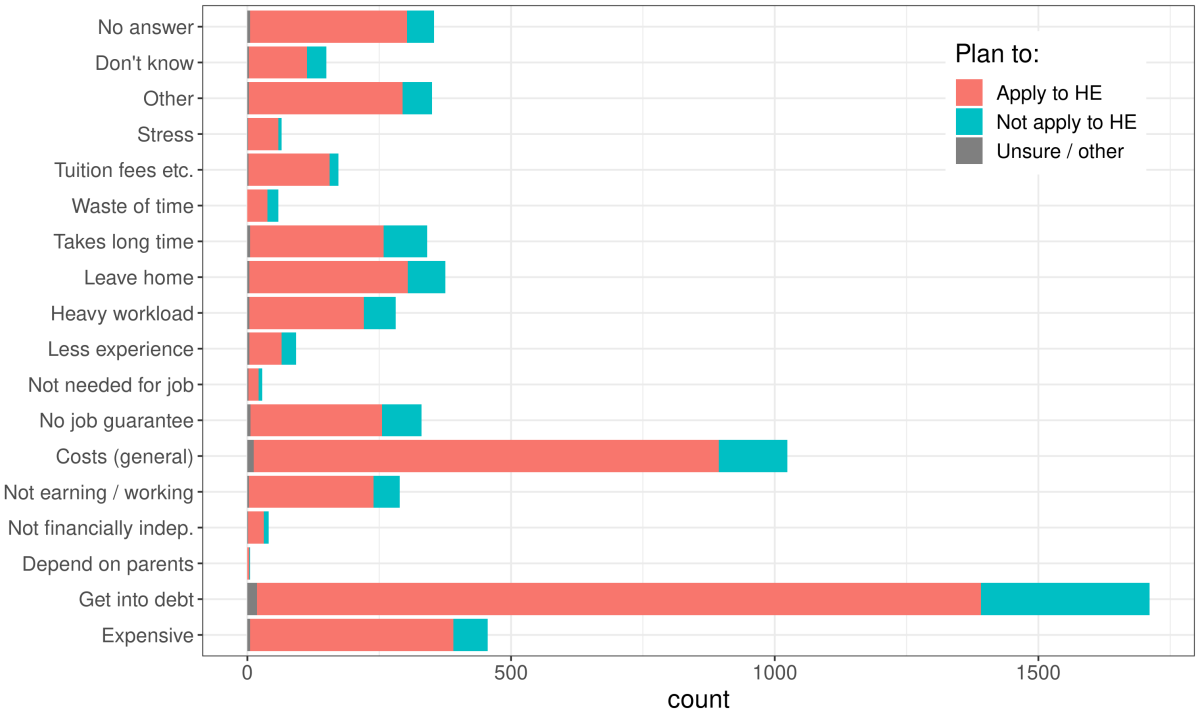
Notes: The "full sample" includes all cohort members who responded in waves 4 *and* 8; the "subsample" is only those who were also asked (and responded to) questions about their expectations regarding univeristy.

Figure 3: Proportion of students who mentioned specific advantages and disadvantages about going to university

(a) Harmonised *advantages*



(b) Harmonised *disadvantages*



Notes: Only students with at least 5 GCSEs at A*–C or equivalent were asked these questions ($N = 4,640$). They were also asked whether they were currently planning to apply to university, which corresponds to the different-colour bar fillings.

estimates. Many of the estimates are sizeable, but they are not estimated with a lot of precision, as evidenced by the large standard errors. Still, many of these variables predict university attendance.

To address the lack of precision in the estimates, we repeat the exercise using an elastic net procedure to select the "best-predictor" explanatory variables among those variables included in the logit regressions. The elastic net combines the penalties of the lasso and ridge variable selection methods to attenuate some of the issues associated with each method alone (Hastie, Tibshirani, and Friedman, 2016). We use the `glmnet` package to implement the elastic net in `R` (Friedman, Hastie, and Tibshirani, 2010). Table B2 in the appendix shows the selected (i.e. non-zero) parameters for different penalties (α) and cross-validated tuning parameters (λ). When using lasso (which corresponds to $\alpha = 1$), the optimal procedure (corresponding to $\lambda_{\min}$) selects many of the advantages and disadvantages, suggesting that these variables contain information relevant to university attendance. The model in section 4 provides the structure and discipline needed to help interpret these parameter estimates and quantify the contributions of different factors to the decision to attend university. Our empirical approach to estimate the model follows in section 5.

4 Theoretical framework

We split the factors young people consider when making educational choices into three categories: future earnings, or pecuniary factors; other financial factors, and everything else, the "non-pecuniary" factors or "psychic costs" (Cunha and Heckman, 2007). Examples of these non-pecuniary factors are: the effort required to gain a place at university; aspects of life at university (social life, studying, leaving home, stress, etc); and aspects of life after university (access to better jobs, graduate "identity", debt).

Utility of university or work. An individual's utility from choosing university ($s = 1$), or work ($s = 0$) is a linear combination of these different factors

$$U_{s,i} = \alpha Y_{s,i} + \theta'_{s,i}\gamma + \epsilon_{s,i} \tag{1}$$

where Y_s represents the pecuniary factors (earnings), $\theta_s \equiv (\theta_s^F, \theta_s^{NP})$ is a vector of financial (θ_s^F) and non-pecuniary (θ_s^{NP}) factors, and ϵ_s is a mean-zero random-utility term, all conditional on choice s.

Decision to attend university. At the time young people make their decision, they do not know the value that many of these outcomes will take, and so form expectations about their utility under each choice, based on the information they hold at that time,

$\mathcal{I}_i$:

$$\mathbb{E}[U_{s,i}|\mathcal{I}_i] = \mathbb{E}[\alpha Y_{s,i} + \theta'_{s,i}\gamma + \epsilon_{s,i}|\mathcal{I}_i] \tag{2}$$

Individuals compare their expected utility of attending university, $U_1^{\mathcal{I}}$,[9] to that of working, $U_0^{\mathcal{I}}$, and choose the option with the higher expected utility. We can write

$$S \equiv \mathbb{1}\{U_1^{\mathcal{I}} - U_0^{\mathcal{I}} > 0\}. \tag{3}$$

This can be rewritten as the difference between expected (pecuniary) outcomes, and expected "costs" of attending university, in the spirit of Roy (1951).

$$S \equiv \begin{cases} 1, & \text{if} \quad \alpha(Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}) - (\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}})'(-\gamma) + \epsilon_1^{\mathcal{I}} - \epsilon_0^{\mathcal{I}} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

This formulation leads naturally to an expression for the probability of attending university, conditional on expected earnings ($Y_s^{\mathcal{I}}$) and "costs" ($\theta_s^{\mathcal{I}}$)

$$\Pr\left(S = 1 | Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}, \theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}}\right) = \Pr\left(\alpha(Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}) - (\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}})'(-\gamma) > \epsilon_0^{\mathcal{I}} - \epsilon_1^{\mathcal{I}}\right) \tag{5}$$

A chief aim of this paper is to estimate the relative importance of the pecuniary and non-pecuniary factors in the decision; i.e. how important is $\alpha(Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}})$ versus $(\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}})'\gamma$ when evaluating this conditional probability.

Earnings, Y_s. We assume the logarithm of earnings is linear-in-parameters, conditional on educational choice, s. Therefore, log-earnings in period t, for individual i with characteristics, $X_{i,t}$, and who chose education s are

$$y_{s,i,t} = X'_{i,t}\beta_{s,t} + v_{s,i,t} \tag{6}$$

where $v_{s,i,t}$ is an unforecastable mean-zero "shock" to earnings. Individuals hold rational expectations about their future earnings, but only possess limited information about the future; the information in their information set, $\mathcal{I}_i$. Then, their expections about future earnings in period t having chosen education s are

$$y_{s,i,t}^{\mathcal{I}} = \mathbb{E}[y_{s,i,t}\,|\,\mathcal{I}_i] \tag{7}$$

$$= \mathbb{E}[X'_{i,t}\beta_{s,t} + v_{s,i,t}\,|\,\mathcal{I}_i]. \tag{8}$$

The additional assumptions about young people's information sets, the process they use to form expectations, and which periods' earnings, $y_{s,t}$, they consider when making their

[9] Employing the shorthand notation $X^{\mathcal{I}} \equiv \mathbb{E}[X|\mathcal{I}_i]$.

educational decisons, Y_s, required to identify the model are discussed in section 5.

5 Empirical strategy

When describing the model in section 4, we alluded to further assumptions necessary to fully identify all of its components. We are now in a position to describe these further assumptions.

5.1 Expectations about (future) earnings

Section 4 presented both a model for wages in period t, conditional on education s, $y_{s,t}$,[10] and a model of how students' expectations about their future earnings conditional on education s, $Y_s^{\mathcal{I}}$,[11] affect their (expected) utility and hence their decision. However, we did not precise the contents of the students' information sets, $\mathcal{I}$, that they use to form these expectations, nor how they discount earnings from different periods.

First, we consider how young people "value" future earnings from different periods when making their decision. Young people only consider their earnings at age 25 (or these are a sufficient statistic for what they consider) when deciding whether to go to university. Therefore,

$$Y_s \equiv y_{s,25} = X_{25}'\beta_{s,25} + v_{s,25}, \tag{9}$$

where variables with a subscript 25 represent realisations of that variable at age twenty-five.

Next, we precise exactly how young people make expectations about Y^s: what is in their information set, and how they use this information to form their expectations. Young people have rational expectations in that they do not make mistakes when forming expectations, but they only possess very limited information about the future—the information in $\mathcal{I}$ reflects their current characteristics, i.e. X_{16}. Put differently, they are very good at predicting mean *realised* earnings among their peers conditional on X_{16}, but they are not very good at predicting their own *future* characteristics (X_{25}). Then $Y_s^{\mathcal{I}} \equiv \mathbb{E}[Y_s \mid X_{16}]$. These assumptions ensure earnings expectations, $Y_s^{\mathcal{I}}$ are identified from (even a subset of) *realised* earnings, and the students' characteristics at seventeen. This is particularly helpful as only one of Y_1 and Y_0 is observed for each individual; I can now construct a counterfactual wage for each student using the data.

[10] The i subscripts are omitted in the interests of cleaner notation.

[11] Recall the shorthand notation $X^{\mathcal{I}} \equiv \mathbb{E}[X|\mathcal{I}_i]$, where $\mathcal{I}$ represents student i's information set when they make their decision about HE.

5.2 Expected other factors "premium"

We use the harmonised responses to open-ended questions about the advantages and disadvantages of going to university, discussed in detail in section 3, to measure the expected other factors premium, $\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}}$. Limited somewhat by the nature of these questions, we assume that individuals either believe there to be no difference in this factor whether they go to university or not, or they believe there will be a difference, which is fixed to be of constant size across all individuals who hold this belief. Therefore for each factor mentioned by *any* student, the component of $\theta_{1,i}^{\mathcal{I}} - \theta_{0,i}^{\mathcal{I}}$ takes one value (normalised to 1) if mentioned by student i, and another value (normalised to 0) if not mentioned.

5.3 Identifying the parameters in the utility function

Recall the probability of attending university, conditional on expectations about earnings and other factors, in the model:

$$\Pr\left(S=1|Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}, \theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}}\right) = \Pr\left(\alpha(Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}) + (\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}})'\gamma > \epsilon_0^{\mathcal{I}} - \epsilon_1^{\mathcal{I}}\right). \tag{10}$$

Identification of α and γ then requires assumptions on the distribution of the error terms, ϵ_1 and ϵ_0. A standard assumption in the discrete-choice literature is that the errors follow a type-I extreme-value distribution, meaning their difference follows a logistic distribution: $(\epsilon_1 - \epsilon_0) \sim \text{Logit}$. These parameters capture the relative contribution of earnings and other factors to young people's utility, and hence in their decision to attend university. Only their ratio is identified by the relative importance of the different factors.

5.4 Estimation

Now the assumptions required to identify the model are clear, we turn to estimation. Although $Y_1^{\mathcal{I}}$ and $Y_0^{\mathcal{I}}$ are identified (and estimated) separately, we only need their difference in the model. The same is true for expectations about other factors, and measurement of their difference, $\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}}$, from data on students' expectations is discussed in this section.

Expected graduate-wage premium, $Y_1^{\mathcal{I}} - Y_0^{\mathcal{I}}$. Under the assumptions in section 5.1, the expected earnings we need are $\mathbb{E}[Y_s|X_{16}]$. Given X_{16} we can use OLS to estimate this conditional expectation, adding interaction terms to capture non-linearities. We assume and estimate the simplest linear conditional expectation for each level of education, with no interactions. We then use the estimated coefficients, $\hat{\beta}_{s,16}$, to obtain estimates $\widehat{Y}_s^{\mathcal{I}} = X_{16}'\hat{\beta}_{s,16}$. The estimated expected graduate-wage premium is simply $\widehat{Y}_1^{\mathcal{I}} - \widehat{Y}_0^{\mathcal{I}} = X_{16}'(\hat{\beta}_{1,16} - \hat{\beta}_{0,16})$. The following covariates are in X_{16}: parents' occupations, parents'

education level, a measure of parental income, the number of A-levels a student is taking, gender, whether high pay is important to them.

Expected premium for other factors, $\theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}}$. We do not need to estimate the expected premiums for other factors as they are simply whether the student mentioned this factor during the survey (see 5.2).

The parameters of the utility function, α and γ. We estimate the parameters of the utility function using logistic regression, (see 5.3).

Distributions of earnings and other factors. A main aim of this paper is to compare and to quantify the roles of earnings, financial, and non-pecuniary factors in the decision to attend university. To do this, we estimate comparable distributions of the different factors using the following strategy: (i) obtain estimates $\hat{\alpha}$, $\hat{\gamma}$, and $\widehat{Y}_1^{\mathcal{I}} - \widehat{Y}_0^{\mathcal{I}}$; (ii) recombine these estimates with the data $(X_{16}, \theta_1^{\mathcal{I}} - \theta_0^{\mathcal{I}})$, to calculate a "utility contribution" for each (type of) factor; (iii) transform these utility values to be equivalent to a percentage-change in earnings; (iv) use a kernel-density estimator to estimate the empirical distributions.

6 Results

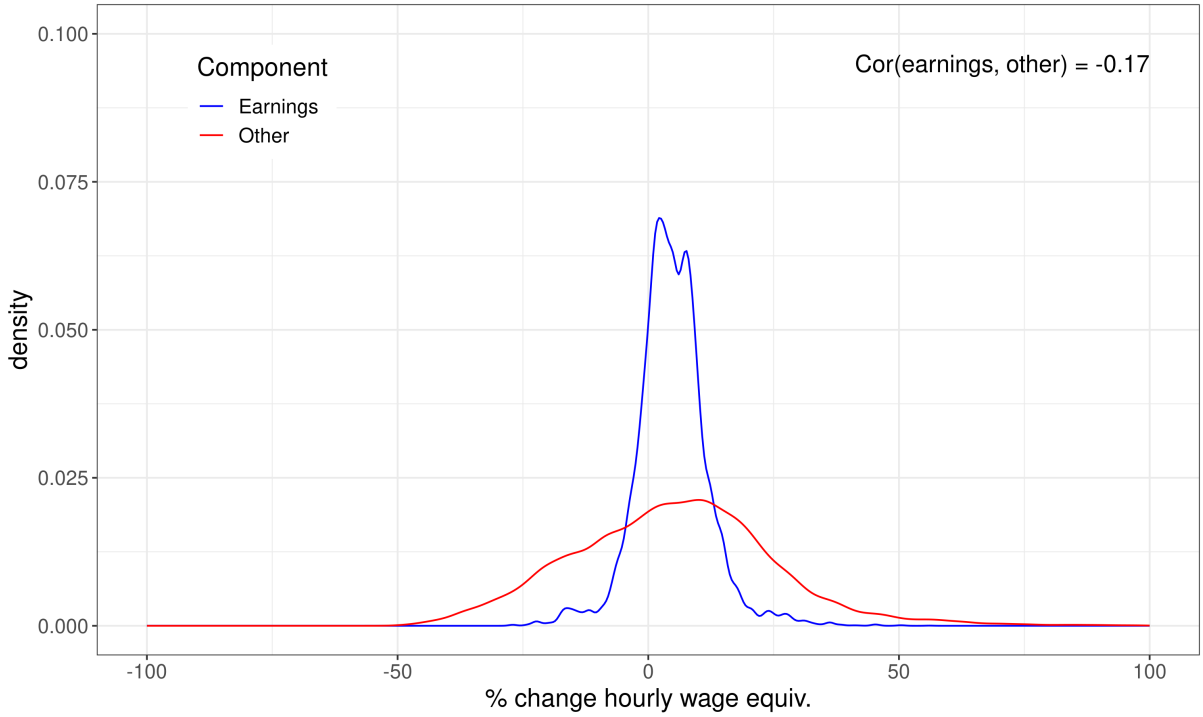
In this section we present and discuss the results of estimating the model discussed in section 3. We focus first on results for the 1990 cohort, initially for the full sample, and then broken down by SES group, using parental income as a measure of SES. These results by SES group allow investigation of the drivers of the SES gap in educational attainment highlighted in figure 2. Finally we present results of estimation on the earlier 1970 cohort, permitting analysis of how factors in the decision have evolved over time.

6.1 1990 cohort: full sample

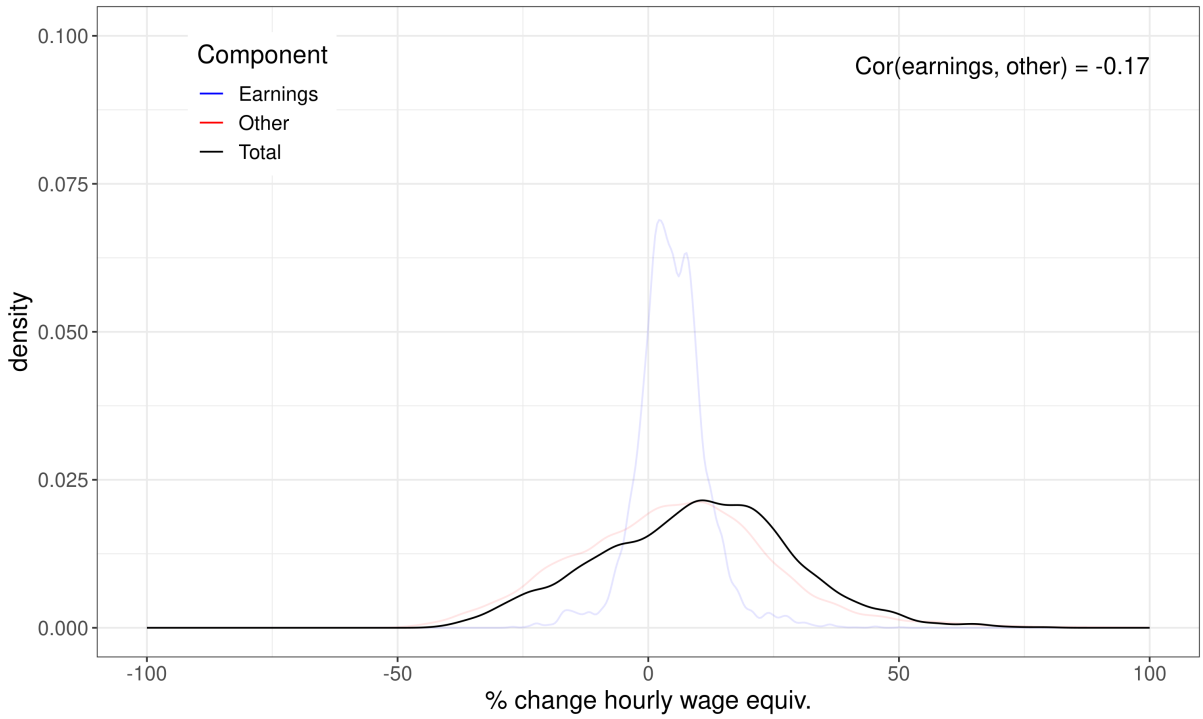
The estimated distributions of earnings (blue line) and other factors (red line) for the full sample of the 1990 cohort are quite striking are in figure 4. The locations of the two distributions are remarkably similar, evidenced by their similar means: 4.75% (pecuniary) and 4.91% (other). The values of young people's earnings and other factors are negatively correlated (-0.17). However, the variance of other factors is much higher (20.0% vs 7.14%). It is chiefly the influence of these other factors that determines whether a student decides to attend university, a role reflected in the similarity between the distribution of all factors in the decision (figure 4b) and the other factors (red line, figure 4a). Consider the following back-of-the-envelope counterfactual exercise to further highlight the importance of other factors.

Figure 4: Distributions of factors in the decision to go to university (1990 cohort)

(a) Earnings versus other factors



(b) Total



Notes: The values of the factors are estimated as described in 5.4. The distributions are then estimated (and plotted) with the kernel density estimator in the `R` package `ggplot2`, using the default Gaussian kernel and bandwidth (Wickham, 2016).

Table 2: Estimated mean, variance, and skewness of the 1990-cohort's factors

	Mean	Std dev.	Skewness
Total	9.38	19.2	34.7
Earnings	4.75	7.14	70.9
Other	4.91	20.0	71.6
Financial	5.41	4.92	125
Non-financial	−2.05	17.8	73.1

Notes: Mean and std dev. are in %Δ wage equivalent.

Table 3: University attendance under counterfactual factor values

Counterfactual	Earnings	Other	University
Data	-	-	0.693
Earnings			
25th percentile	0.87	-	0.619
75th percentile	8.58	-	0.748
Other			
25th percentile	-	−8.68	0.240
75th percentile	-	17.2	0.995

Notes: Earnings and other factors are in %Δ wage equivalent. University is the fraction who attend.

Counterfactual exercise. We fix the value of one type of factor for all individuals, and then calculate how many people would still decide to attend university. The results are in table 3. Varying the expected graduate-wage premium between the 25th or 75th percentile has a limited effect on university attendance, with only 9 percentage points (*pp*) fewer people attending university at the 25th percentile, and 5*pp* more people attending at the 75th. The effects of the same exercise with other factors are much larger. If everyone in the sample had other factors equal to the 25th percentile, only 24% of people would attend university—over 65*pp* fewer than did actually attend. Meanwhile, assigning everyone other factors equal to the 75th percentile results in over 99% of people attending university. Although this back-of-the-envelope calculation abstracts from potential equilibrium effects, it demonstrates the importance of non-earnings factors versus earnings.

Separating financial from other (non-pecuniary) factors. So far financial and non-pecuniary factors were grouped together in our analysis. We separate financial factors from non-pecuniary factors in figure 5. The young people in our sample mentioned a number of *financial* factors: "tuition fees", "costs (general)", "not earning / working", "not financially independent", "depend on parents", "get into debt", "expensive". Previous research has highlighted the psychological burden of financial concerns (Gathergood, 2012),

and financial factors are often offered as an explanation for why more young people do not attend university (Keane and Wolpin, 2001). Therefore quantifying their role here is important. By removing financial factors, we also isolate the truly non-pecuniary factors—or "psychic costs". Figure 5a compares financial factors to earnings, which appear to share remarkably similar distributions. These similarities are also apparent in table 2, though there they are perhaps less striking. Moreover it is clear from figure 5b that financial factors are not driving the variance of the other factors.

6.2 Splitting the 1990 cohort by socio-economic status

In this section we present the factor distributions *conditional on socio-economic status*. We use parental earnings at age sixteen as a measure of socio-economic status (SES), in part to ensure our analysis is comparable to previous work on this subject (see e.g. Blanden and Machin (2004)). Comparing the factor distributions across SES allows us to quantify the relative contributions of earnings and other factors to the SES-gap in university attendance (see figure 2 and table 1).

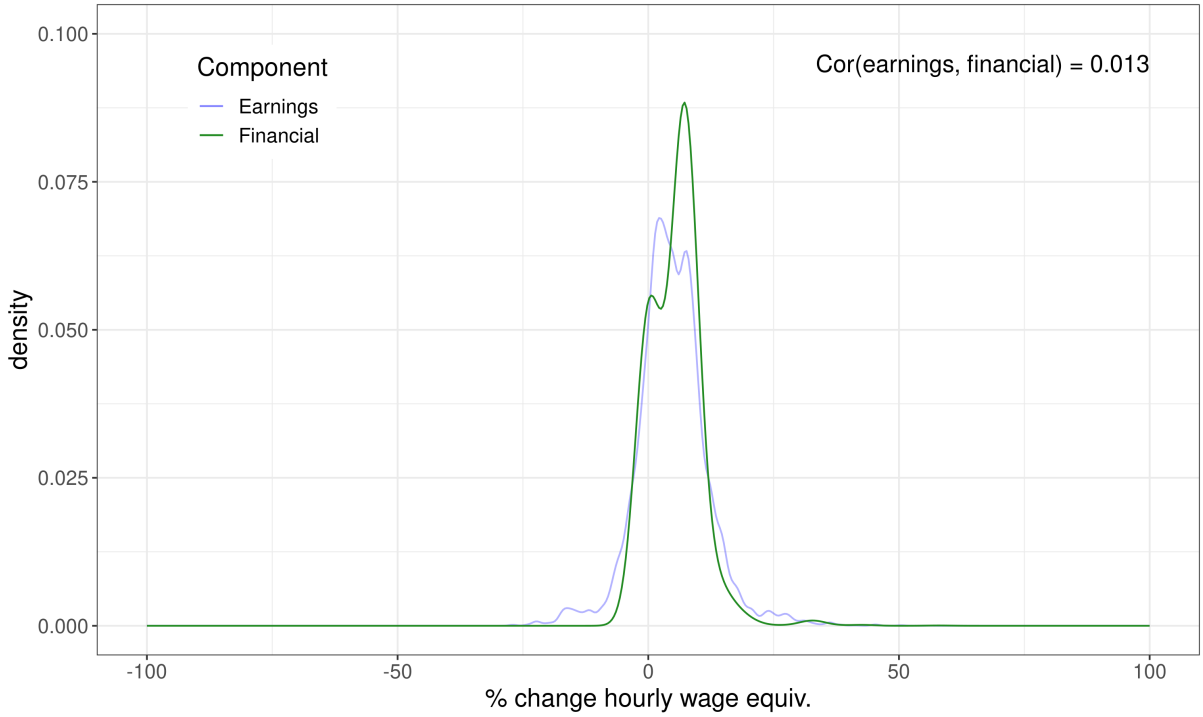
Figure 6 shows the distributions of earnings (left column) and other (right column) factors, for those with parents in the bottom 20% (top row), middle 60% (middle row), and top 20% (bottom row) of the earnings distribution. Focusing first on earnings (left column, figure 6), the distributions of factors across the three groups are similarly located, though the means are decreasing in parental income (table 6g). The opposite is true for other factors (right column, figure 6): the distributions of factors across the three groups clearly occupy different locations, and their means are strongly increasing in parental income. The mean other factors in the bottom SES-group are slightly negative (-0.74) while they are over 10 for the top SES-group. The SES-gap in educational attainment is entirely driven by other factors in our analysis.

6.3 Changes in factors across cohorts

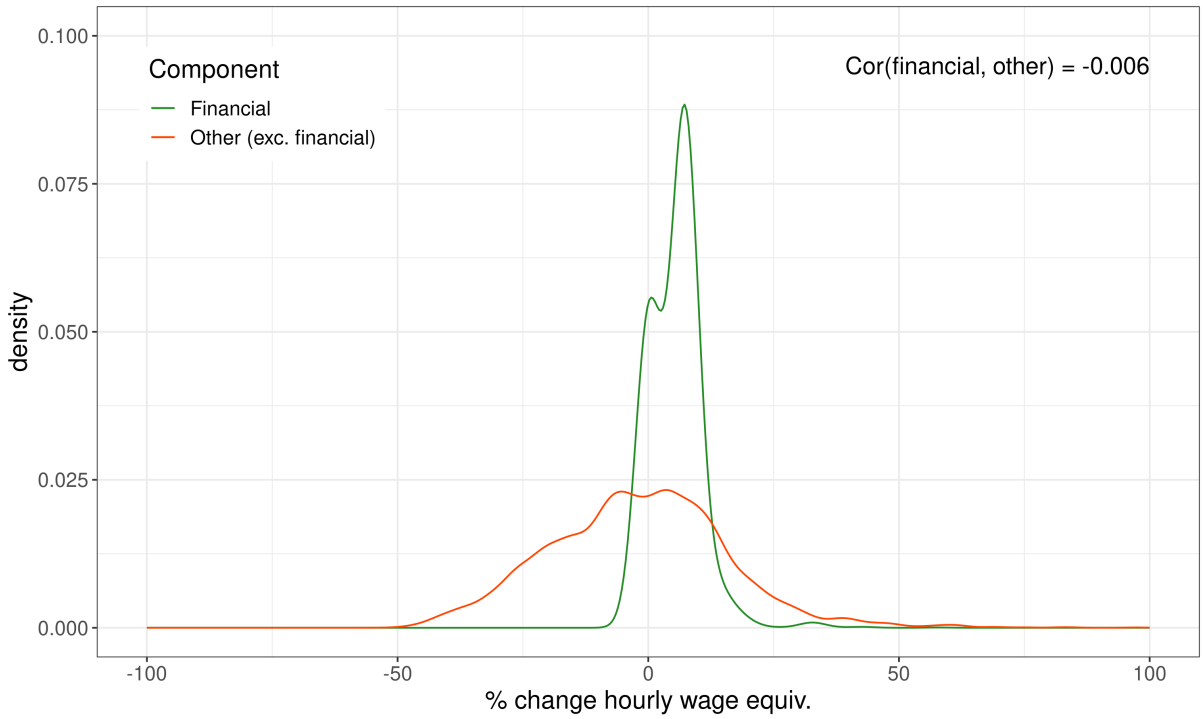
In an effort to shed light on what drove more and more people to attend university in England in recent decades (figure 2), we re-estimate our model on data from a cohort born in 1990 (see section 3). The results are striking. Figure 7 presents the estimated distributions of factors for the two cohorts, again split into earnings and other factors. The mean graduate-wage premium actually *decreased* on average between the two cohorts, a change accompanied by a reduction in variance. Meanwhile, other factors *increased* significantly on average over this period, and their variance also increased slightly. Taken together, these results mean the large increase in higher education attainment between the two cohorts (see table 4) was entirely due to an increase in expectations about factors other than earnings.

Figure 5: Decomposing other factors into financial and non-pecuniary

(a) Earnings versus financial factors



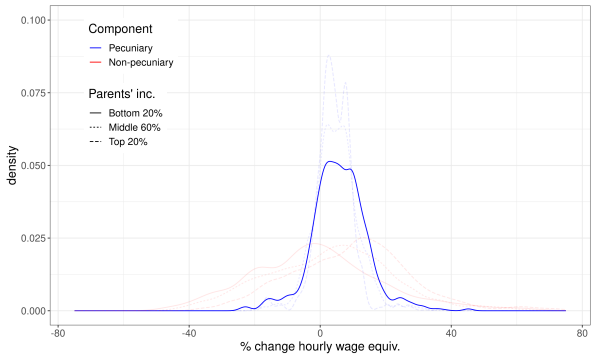
(b) Financial versus other (non-pecuniary) factors



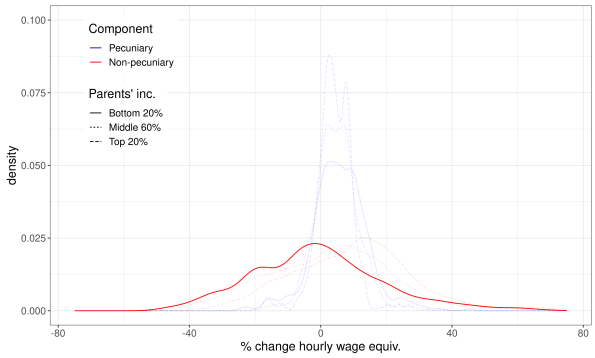
Notes: The values of the factors are estimated as described in 5.4. The distributions are then estimated (and plotted) with the kernel density estimator in the `R` package `ggplot2`, using the default Gaussian kernel and bandwidth (Wickham, 2016). Financial factors are "tuition fees", "costs (general)", "not earning / working", "not financially independent", "depend on parents", "get into debt", "expensive".

Figure 6: Comparing factor distributions by parental income (SES)

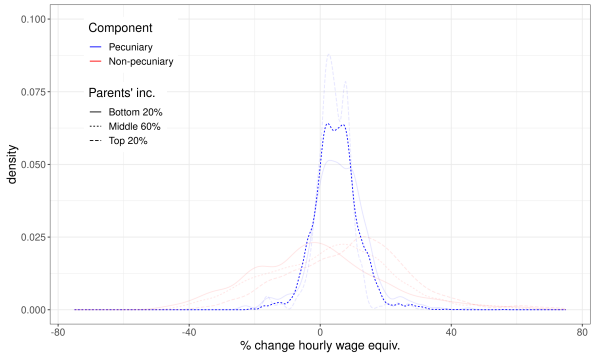
(a) Earnings, bottom 20%



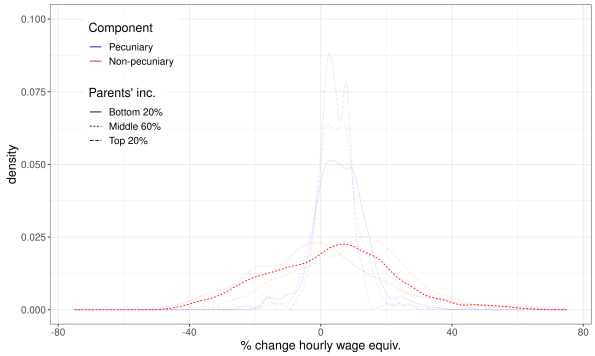
(b) Other, bottom 20%



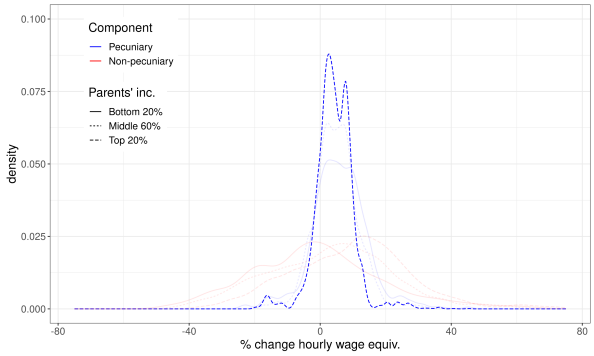
(c) Earnings, middle 60%



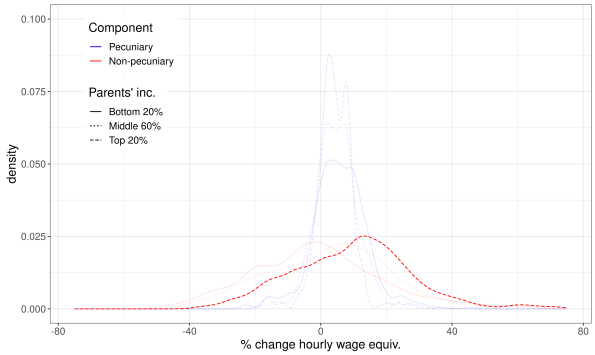
(d) Other, middle 60%



(e) Earnings, top 20%



(f) Other, top 20%



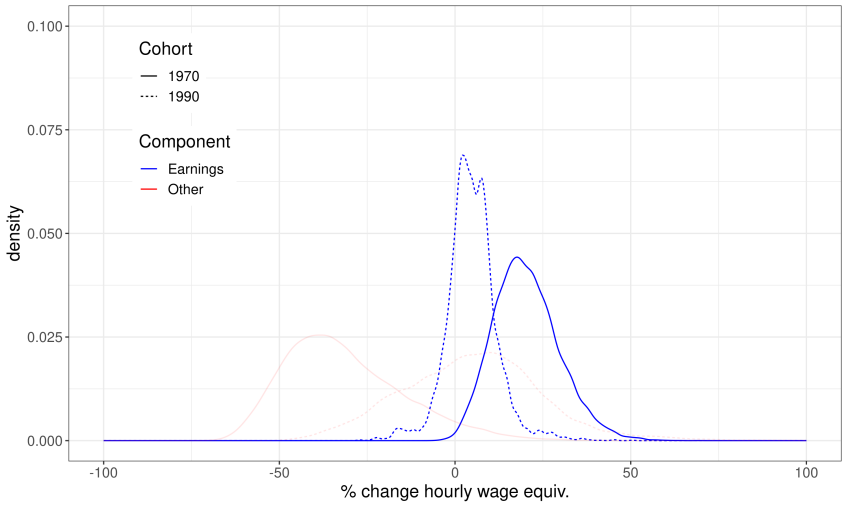
(g) Summary statistics for the distributions in panels (a)–(f)

	Mean	Std dev.	Skewness
Earnings			
Bottom 20%	5.92	8.64	19.0
Middle 60%	4.94	6.82	38.3
Top 20%	4.32	5.92	28.5
Other			
Bottom 20%	−0.74	20.4	47.8
Middle 60%	3.22	19.5	26.2
Top 20%	10.2	19.5	62.6

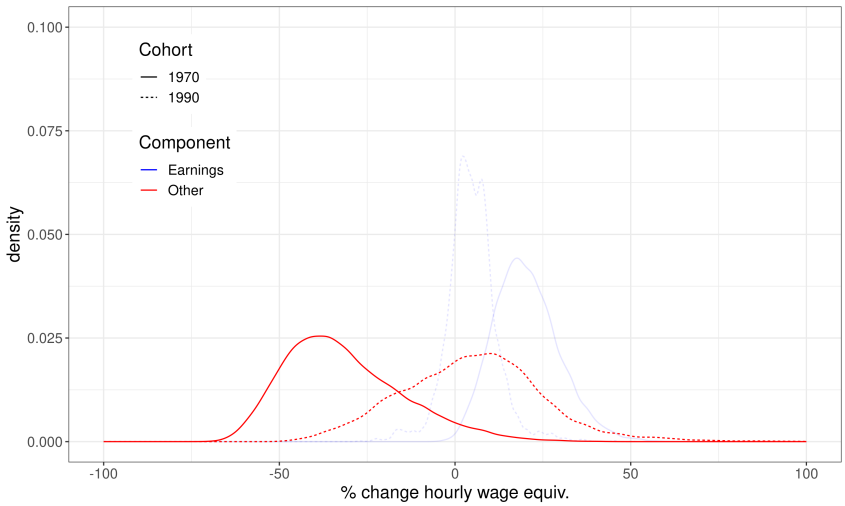
Notes: The values of the factors are estimated as described in 5.4. The distributions are then estimated (and plotted) with the kernel density estimator in the `R` package `ggplot2`, using the default Gaussian kernel and bandwidth (Wickham, 2016). The mean and standard deviations in panel (g) are in %Δ wage equivalent.

Figure 7: Changes in distributions of factors between cohorts (1970–1990)

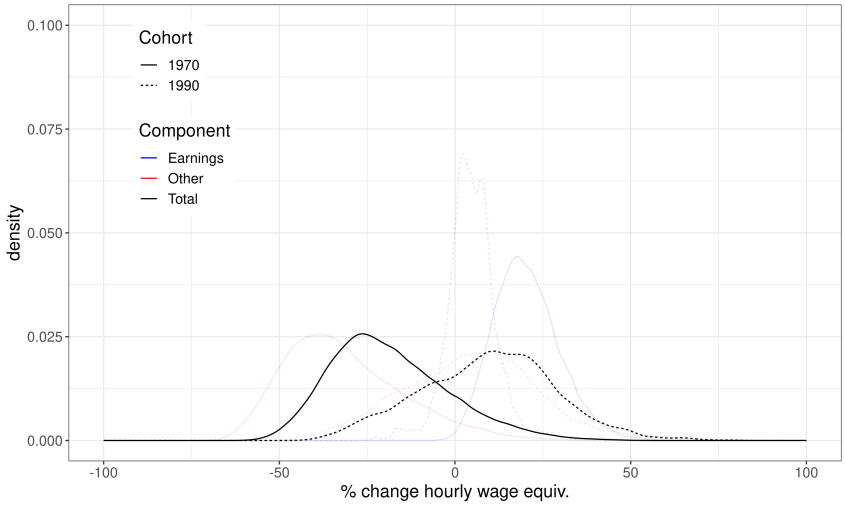
(a) Earnings



(b) Other



(c) Total



Source: The 1970 distributions are estimated using the BCS data, and 1990 using Next Steps.

Notes: The values of the factors are estimated as described in 5.4. The distributions are then estimated (and plotted) with the kernel density estimator in the `R` package `ggplot2`, using the default Gaussian kernel and bandwidth (Wickham, 2016).

Table 4: Comparison of summary statistics between cohorts

Cohort:	BCS (1970)	LSYPE (1990)	Change 1970–1990
Degree	0.14	0.69	+0.6
Pecuniary			
Mean	20.4	4.75	−16.7
Std dev.	9.28	7.14	−2.1
Non-pecuniary			
Mean	−31.5	4.91	+36.4
Std dev.	17.0	20.0	+3.0

Notes: The mean and standard deviations in panel (g) are in %Δ wage equivalent.

7 Allowing for unobserved heterogeneity

The results so far are quite striking: expectations about non-pecuniary factors are much more important in determining whether someone attends university than expectations about earnings. However, to reach these results we relied upon some strong assumptions, particularly concerning how students form their earnings expectations, and which periods' earnings they consider. In this section, we relax some of these assumptions, first specifying an updated model which includes a key new feature for earnings expectations: students hold information at 16 which they use to forecast their future earnings which the econometrician does not observe ("unobserved heterogeneity"). We capture this unobserved heterogeneity with latent types, which once estimated we can treat like an additional *observed* characteristics in my model. Therefore, the chief innovation of this section concerns identifying and estimating this *unobserved* component of earnings, which requires a novel methodology that we developed in Cassagneau-Francis, Gary-Bobo, Pernaudet, and Robin (2020, henceforth CGPR).

7.1 The model

The utility function and decision process remain unchanged from the model exposed in section 4. Therefore we will only briefly restate them here, and invite the reader to refer back to previous sections for more detail.

Utility of university or work. An individual's utility from choosing university ($s=1$), or work ($s=0$) is a linear combination of pecuniary and non-pecuniary factors

$$U_{s,i} = \alpha Y_{s,i} + \theta'_{s,i}\gamma + \epsilon_{s,i}. \tag{11}$$

where Y_s represents the pecuniary factors (earnings), θ_s is a vector of financial and non-pecuniary factors, and ϵ_s is a mean-zero random-utility term, all conditional on choice s.

Decision to attend university. Individuals compare their expected utility of attending university, $U_1^{\mathcal{I}}$,[12] to that of working, $U_0^{\mathcal{I}}$, and choose the option with the higher expected utility. We can write

$$S \equiv \mathbb{1}\{U_1^{\mathcal{I}} - U_0^{\mathcal{I}} > 0\}. \tag{12}$$

The key innovations of this model over the one in section 4 concern earnings, Y_s, and students' information sets, $\mathcal{I}$. In particular, there is an individual-specific component of earnings that individuals know about at the time they make their decision (it is in their information set, $\mathcal{I}$), but which is unobserved by the econometrician. Crucially, this individual-specific component of earnings is also important for their academic performance—and hence their test scores at sixteen. In practice, we assume that this individual-specific component does not vary continuously across individuals, but instead takes one of K discrete values, which are captured by an individual's "type".

Individual types. We assume that we can classify individuals into K groups or types, denoted $k \in \{1, ..., K\}$. Their type determines (in part) their wages, whether they attend university, and their test scores and preferences at sixteen. Therefore, these variables also contain information on an individual's type that we can exploit.

Test scores at sixteen. We denote the probability density function (PDF) of observing a test score τ conditional on type, k, and observables at sixteen, X_{16} by $g(\tau|k, X_{16})$.

Earnings. Individuals' log-earnings in period t are drawn from a distribution which depends on their type, k, their (observed) characteristics in t, X_t, and their education, s, which we write as $f(y_t|k, X_t, s)$. Therefore, we can write log-earnings at age 25 for individual i, who is of type k, as

$$y_{i,s,t} = \alpha_{s,k} + X'_{i,t}\beta_{s,t} + u_{s,i,t}. \tag{13}$$

Relative to the model for earnings in section 4 there are two key differences: (i) the constant term $\alpha_{s,k}$ which previously did not vary with type (and was implicitly a component of $\beta_{s,t}$);[13] and (ii) the error term, $u_{s,i,t}$, which still represents a mean-zero random productivity shock, though drawn from a distribution whose variance depends on type, k, as well as schooling, s.

[12] Employing the shorthand notation $X^{\mathcal{I}} \equiv \mathbb{E}[X|\mathcal{I}_i]$. Recall that $\mathcal{I}$ is the individual's information set when they make their decision (at age sixteen).

[13] Alternatively the previous model was implicitly assumed to include sufficient observed variation through X_t to capture variation in earnings due to latent types. That assumption is relaxed here.

Information set, $\mathcal{I}$. Individuals form expectations in the same way as before, with one key difference: where before their information set contained only their cohort's mean returns conditional on observed characteristics at sixteen, they now also know their type, k, and the type-conditional mean returns among their peers (cohort). In the previous notation, we now have $Y_s^{\mathcal{I}} \equiv \mathbb{E}[Y_s|X_{16}, k]$.

7.2 Identification

As we only observe earnings at age twenty-five for the main sample, we focus on identifying the model from data from this period. We also use slightly different notation in this section for clarity of exposition, and will attempt to point out where this coincides with notation elsewhere in the paper.

Under the latent types assumption, the complete likelihood of a given data point is

$$p(y, s, \boldsymbol{z}, \boldsymbol{q}) = \sum_{k=1}^{K} \pi(k, \boldsymbol{z}, s) f(y|k, \boldsymbol{q}, s) \tag{14}$$

where $\boldsymbol{z}$ contains observable characteristics known to the individual at 16 ($\boldsymbol{z} \supset \{X_{16}, \theta_s\}$), $\boldsymbol{q}$ ($\equiv X_{25}$) contains observable characteristics that determine wages at twenty-five but were (perhaps) *unknown* at 16, $\pi(k, \boldsymbol{z}, s)$ is the probability of being type k, with characteristics $\boldsymbol{z}$ and choosing schooling s, and $f(y|k, \boldsymbol{q}, s)$ is the distribution of earnings at twenty-five conditional on type, observed characteristics (at twenty-five), and schooling.

Consider $\tilde{y}_s \equiv y_s - \boldsymbol{q}\beta_s$; an earnings residual "cleansed" of the effects of $\boldsymbol{q}$. For simplicity, also assume that we only have a single, binary instrument z. The new likelihood without $\boldsymbol{q}$ is,

$$p(\tilde{y}, s, z) = \sum_{k=1}^{K} \pi(k, z, s) f(\tilde{y}|k, s) \tag{15}$$

7.2.1 Identifying $\pi(k, z, s)$ and $f(\tilde{y}|k, s)$

In CGPR, we show that to non-parametrically identify the elements on the right-hand side of equation (15) we need an additional continuous measurement that depends on latent types, but not on treatment.[14] A natural solution would be to use earnings from before "treatment" as we do in our application in CGPR, where the treatment is formal training. Here, as treatment is university we cannot use "pre-treatment" wages for most individuals. However, if latent types capture in some sense both innate ability / personality and the environment in which a person was raised, grades in school (or "test scores" in section 7.1) contains information on these types. We have the continuous observation we need.

[14] In fact types are identified even with all measurements dependent on treatment, though it requires an additional assumption to ensure types are consistent across treatment groups.

We denote test score by τ. The probability of observing an individual with instrument z, schooling, s, test score, τ and wage at twenty-five, $\tilde{y}$, is

$$p(\tau, \tilde{y}, s, z) = \sum_{k=1}^{K} \pi(k, z, s) g(\tau|k) f(\tilde{y}|k, s). \tag{16}$$

Note the assumption that conditional on type, test scores are independent of whether someone goes to university. This seems reasonable given our assumptions about what types capture. We refer the reader to CGPR for a detailed identification proof, and we state only the necessary assumptions and provide some intuition in the appendix C.

7.3 Estimation

To estimate the model, we impose parametric forms on the functions f and g.[15] Specifically, we assume log-earnings and test scores are normally distributed, so that

$$g(\tau|k) = \frac{1}{\sigma_\nu} \varphi\left(\frac{\tau - \mu_k}{\sigma_\nu}\right) \tag{17}$$

$$f(y|k, X_{25}, s) = \frac{1}{y\sigma_\varepsilon} \varphi\left(\frac{\ln y - \alpha_{ks} - X_{25}'\beta_s}{\sigma_\varepsilon}\right) \tag{18}$$

where $\varphi(\cdot)$ is the standard Gaussian density function. We use the posterior probabilities to estimate $\pi(k, \boldsymbol{z}, s)$, so that

$$\hat{\pi}(k, \boldsymbol{z}, s) = \frac{1}{N} \sum_{i: z_i = z, s_i = s} p_i(k) \tag{19}$$

Under these assumptions, we can obtain an estimate of the model's parameters via maximum likelihood aided by the expectation-maximisation (EM) algorithm (Dempster, Laird, and Rubin, 1977). We provide further details of our estimation strategy and EM algorithm in appendix D.

7.3.1 Adding the unobserved component to the decision model

The outcome of this estimation strategy is the complete set of parameters of the wage model specified in section 7.1, plus posterior probabilities $p_i(k)$ for each individual. We can then proceed in one of two ways: either assign each individual their type corresponding to their highest posterior probability, $k_i = \arg\max_k p_i(k)$; or assign to each individual an individual specific intercept, α_i, which is a weighted sum of their posterior probabilities and the type-specific intercepts, α_k, i.e. $\alpha_{is} = \sum_k p_i(k)\alpha_{ks}$.

Estimating the model exposed in this section is currently in progress.

[15] We also return to the notation used prior to section 7.2, except $y \equiv y_{25}$, and $\boldsymbol{z} = (X_{16}, \theta_1 - \theta_0)$.

Once we obtain these estimates we can include them in the decision model *as if* they were another observed characteristic. We include α_{is} or α_k as a component in X_{16} and proceed with estimation as in section 5.4.

8 Conclusion

In this paper we specify and estimate a model of educational choice, that specifically includes expectations about earnings and other, financial and non-pecuniary, factors. We exploit data on a cohort born at the end of the 1980s which features data on realised earnings and expectations about the non-pecuniary costs and benefits of going to university. Our findings add support to the notion that individuals are not strict income maximisers when they make educational choices. We find that non-pecuniary expectations are able to explain most of the variation across individuals that causes some people to attend university and others to not, with the expected graduate-earnings premium playing a minor role. Splitting the sample by parental income (a measure of socio-economic status), we find that differences in factors other than earnings across socio-economic groups are responsible for the "SES gap" in educational attainment. Finally, comparing the roles of pecuniary and other factors in educational decisions across a period of significant growth in higher education attainment and increased financial costs, we find that the expected graduate premium fell slightly, suggesting increases in the value of non-pecuniary factors drove the expansion in attainment.

Our findings join a growing body of evidence that non-pecuniary factors play a key role in educational decisions. Results from the improved model of expected earnings will ensure that the results presented so far do not underestimate the role of pecuniary factors. Careful work decomposing the non-pecuniary factors into meaningful components is another vital next step.

References

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020): "Ex Ante Returns and Occupational Choice," *Journal of Political Economy*, 128, 4475–4522.

Arcidiacono, P. and J. B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71, 933–946, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00431.

Blanden, J. and S. Machin (2004): "Educational Inequality and the Expansion of UK Higher Education," *Scottish Journal of Political Economy*, 51, 230–249.

Blundell, R., D. Green, and W. Jin (2018): "The UK Education Expansion and Technological Change," 71.

Blundell, R., D. A. Green, and W. Jin (2021): "The UK as a Technological Follower: Higher Education Expansion, Technological Adoption, and the Labour Market," *Review of Economic Studies*, 81.

Boneva, T. and C. Rauh (2020): "Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns," *HCEO Working Paper*, 83.

Cassagneau-Francis, O., R. J. Gary-Bobo, J. Pernaudet, and J.-M. Robin (2020): "A Nonparametric Finite Mixture Approach to Difference-in-Difference Estimation, with an Application to Professional Training and Wages," *Working Paper*.

Crawford, C. and W. M. Jin (2014): "Payback time? Student debt and loan repayments: what will the 2012 reforms mean for graduates?" Tech. rep., Institute for Fiscal Studies.

Cunha, F. and J. J. Heckman (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14, 870–893.

Cunha, F., J. J. Heckman, and S. Navarro (2004): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57, 191–261.

Dearden, L., E. Fitzsimons, and A. Goodman (2005): "Higher education funding policy: a guide to the debate," .

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, pp. 1–38, publisher: Wiley for the Royal Statistical Society.

Devereux, P. J. and W. Fan (2011): "Earnings returns to the British education expansion," *Economics of Education Review*, 30, 1153–1166.

D'Haultfoeuille, X. and A. Maurel (2013): "Inference on an extended Roy model, with an application to schooling decisions in France," *Journal of Econometrics*, 174, 95–106.

Friedman, J., T. Hastie, and R. Tibshirani (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22.

Gathergood, J. (2012): "Debt and Depression: Causal Links and Social Norm Effects," *The Economic Journal*, 122, 1094–1114, publisher: Oxford Academic.

Gong, Y., T. Stinebrickner, and R. Stinebrickner (2019): "Uncertainty about future income: Initial beliefs and resolution during college," *Quantitative Economics*, 10, 607–641, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE954.

Hastie, T., R. Tibshirani, and J. Friedman (2016): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, New York, NY: Springer, 2nd edition ed.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018): "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126, S197–S246, publisher: The University of Chicago Press.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, Elsevier, vol. 1, 307–458.

Keane, M. P. and K. I. Wolpin (2001): "The effect of parental transfers and borrowing constraints on educational attainment," *International Economic Review*, 42, 1051–1103.

OECD (2018): *Equity in Education: Breaking Down Barriers to Social Mobility*, PISA, OECD.

Oreopoulos, P. and U. Petronijevic (2013): "Making College Worth It: A Review of the Returns to Higher Education," *The Future of Children*, 23, 41–65.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

University College London, UCL Institute of Education, Centre for Longitudinal Studies (2018): "Next Steps: Sweeps 1–8. 2004–2016. [data collection]." .

Walker, I. and Y. Zhu (2008): "The College Wage Premium and the Expansion of Higher Education in the UK," *The Scandinavian Journal of Economics*, 110, 695–709.

Wickham, H. (2016): *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Wiswall, M. and B. Zafar (2015): "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 82, 791–824.

Woodin, T., G. McCulloch, and S. Cowan (2013): *Secondary Education and the Raising of the School-Leaving Age: Coming of Age?*, Secondary Education in a Changing World, Palgrave Macmillan US.

A Institutional context of HE in England

In this section we discuss the organisation of higher education in England. Schooling is compulsory up to the age of sixteen in the UK, and has been since 1972 (Woodin, McCulloch, and Cowan, 2013). Figure A1 presents the time-line of decisions and exams that students (generally) must take to secure a place at university. Two key decisions are: the application to continue on to further education ("sixth form") in the final year of secondary school; and the university application in the final year of sixth form. The main data source follows individuals through secondary school and beyond, from the age of 14 until 19. However, in this paper we will focus exclusively on the decision to attend university and treat the outcome of the decision to continue to sixth form as a predetermined characteristic. Estimating a dynamic discrete-choice model to exploit more of the data is an interesting avenue we hope to explore in future work.

University application process. The UK university application system is quite unique in many ways, and is worthy of study in its own right. Students apply through a centralised system, the "Universities and Colleges Admissions Service" (UCAS)[16] in the autumn of their final year of sixth form. Students can apply for up to five places, where each "place" is a *university-subject pair*. The application consists of a personal statement written by the student, predicted A-levels grades from their teachers, and past national-exam results. These are common across all applications, so students cannot tailor their personal statement to different subjects or institutions.[17] Students then receive *conditional* offers or are rejected from each place they applied, and must select two of their offers: a first choice and a back-up option. The offers made to students in sixth form are (almost exclusively) conditional on their future grades, so for example may require a student sitting 3 A-levels to achieve AAB, with one A in chemistry. The back-up option allows the student to aim high with their first choice, and still have a place somewhere if they fail to achieve those grades. Students sit their A-levels knowing their required grades for each place, and are automatically accepted at their first choice if they achieve the required grade, at their second if they miss the requirement for their first choice, and nowhere if they do not meet either requirement.[18]

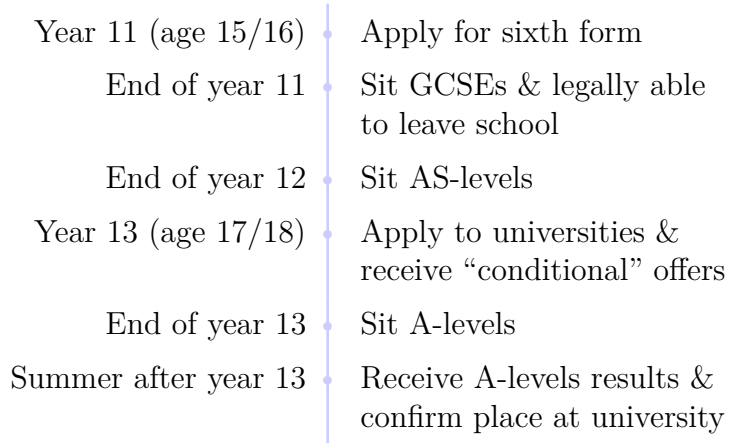
The funding of higher education. Universities in the UK are privately run, but receive state funding *and* are regulated by government over the fees they can charge their students. Tuition fees were first introduced for UK students at UK universities in 1998. Prior to this, universities could not charge fees for tuition. There was also a

[16] Universities and colleges are different entities in the UK, and the names are not used interchangeably, unlike in the US.

[17] This is an implicit barrier which stops people applying to vastly different subjects.

[18] There is a mechanism to allocate students who missed their offers on both their first- and second-choices to places at universities who remain unfilled called "Clearing".

Figure A1: Timeline of educational decisions (1990 cohort)



system of grants and loans in place to cover living costs. In 1998 a means-tested fee was introduced, with the students from the most privileged backgrounds paying £1,000 per year in tuition fees. The poorest students were entitled to a 25% reduction. The situation changed again in 2006, with the introduction of so-called "top-up" fees, which could be set by each university up to a maximum of £3,000.[19] Alongside these fees, the government introduced a comprehensive system of loans and grants to cover both tuition fees and living costs ("maintenance"). Grants and some loans were means-tested, but all students could borrow the full fee, plus some extra for maintenance. The repayment schedule of the loans was made income contingent, meaning that no repayments were required until a graduate earned over a threshold amount, and repayments were set at a percentage of all earnings over this threshold. Therefore, not only does attending university affect the earnings that someone might expect to receive, but their (expected) future earnings will affect how much they expect to pay for their degree, a key feature to capture in the model.

Tuition fees, student loans and maintenance grants The funding of higher education in the UK has changed frequently in recent years moving from a model of direct government funding prior to 1998, to a model with increasingly higher tuition fees alongside a system of government-subsidised loans and grants (see Table 2.1 in Crawford and Jin (2014) for a summary of some of these changes). The majority of the individuals in the main cohort we use left sixth-form in 2007, so they would have experienced the system under reforms that came into force in 2006, henceforth the "2006 reforms". The key features of the system under the 2006 reforms are summarised in table A1.

Student debt levels on graduation. Dearden, Fitzsimons, and Goodman (2005) calculate expected debt levels for a student entering university in 2006/7 (i.e. under the

[19]This maximum fee is set currently at slightly over £9,000, though the increase occurred after the relevant period for the analysis in this paper (in 2012).

Table A1: Details of fees, loans and grants available under the 2006 reforms

Measures	Details
Tuition fees	• Set by university, up to £3,000 p.a. • payable by ALL students
Grants	• Means-tested up to £2,700 p.a. • Tapered to zero at £33,560.
Loans	
Fees	• Equal to fees charged by university. • Available to ALL students.
Maintenance	• £3,555 p.a. if household income <£26,000. • Loan increases from £3,555 p.a. incrementally • Up to £4,405 p.a. if family income between £26,000 and £33,560. • Tapered down to £3,305 at £44,000.
Repayment	• 9% of income above £15,950 (threshold rises with inflation). • State-subsidised loans, zero-real interest rate. • Debt forgiven after 25 years.

Source: Crawford and Jin (2014)

Table A2: Expected debt on graduation (maximum loans under 2006 reforms)

Parental income	Debt on graduation	Share in sample
Low (<£15,970 p.a.)	£19,340	0.20
Middle (∼£25k p.a.)	£19,340	0.09
Upper middle (∼£30k p.a.)	£21,440	0.22
High (>£44k p.a.)	£18,670	0.31
Missing income info.	-	0.18

Source: Dearden et al. (2005) (debt figures) and author's calculations.

first year of the 2006 reforms). Their calculated expected debts are in table A2, along with the share of individuals in each category in the Next Steps cohort. The information in tables A1 and A2 show that although the sticker price of education in the UK was quite high, loans were available to all suggesting credit constraints are not an issue in the UK context. In addition the (maximum) debt burden faced by students appears to be relatively constant across socio-economic groups (though of course the psychological effects of this debt may still vary).

B Data appendix

B.1 Next Steps

Other information collected in wave four

In addition to the data on expectations collected in wave 4, I also use information on family background and schooling up to age sixteen. I use detailed information on parental earnings to estimate a measure of socio-economic status (SES), based on the quintiles of parental earnings (I also use an alternative definition based on means-tested grant eligibility, again calculated from parental earnings). I include information on parents' occupations, ethnicity, education, and income in the model, as well as (limited) information on ability[20] (number of A-levels being taken), and gender. Table 1 presents descriptive statistics for these variables.

B.2 LSYPE wave eight (age twenty-five)

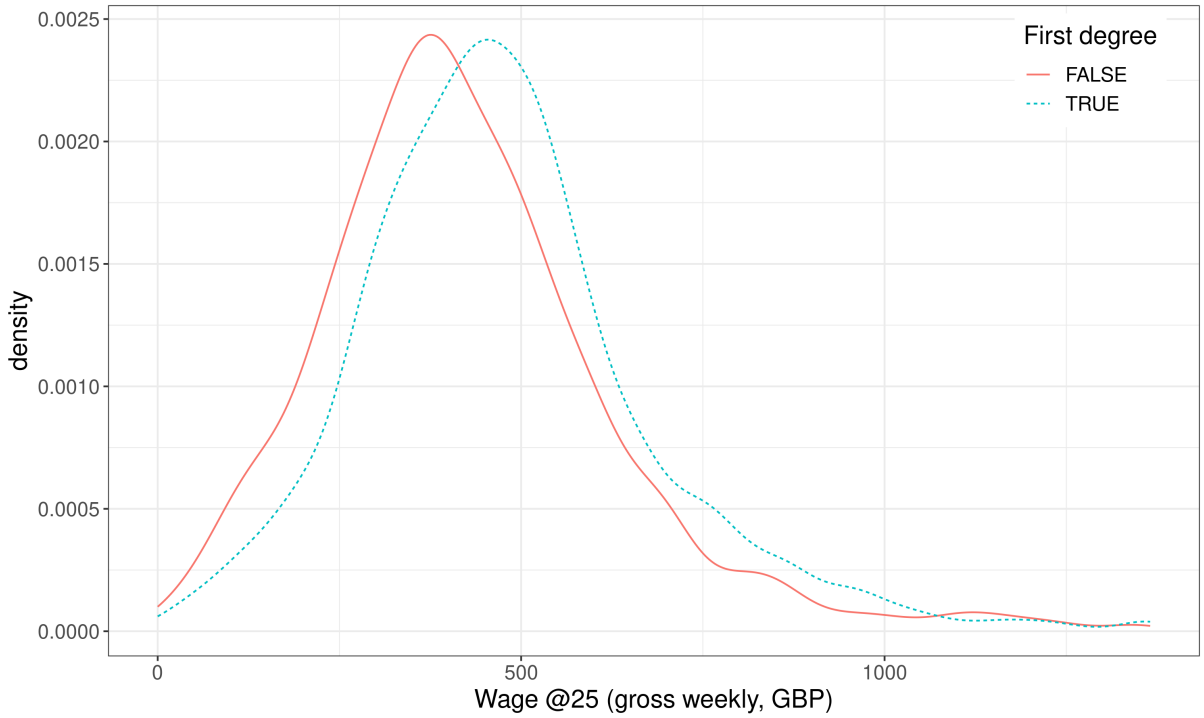
The other key wave of Next Steps for my analysis is the eighth, when the cohort members are aged 25. At this point the majority are working (or at least have worked at some point), and most of those who attend university have completed their degrees.

Degree attainment. The cohort members are asked about any qualifications they have achieved since the last interview (wave seven, five years previously), including whether they hold an undergraduate degree. Table 1 shows information on the proportion of cohort members who hold a degree at 25, including the proportion who attended a member of the Russell Group (a "club" of prestigious research universities in the UK). I also break down degree attainment by SES group (parental income quintiles) in table 1. All these statistics are shown for all respondents to waves 4 and 8, and for the subsample who answered questions about university. Nearly 70% of the analysis subsample hold a degree by the time they are 25, though there is still substantial variation across socio-economic groups reflecting the patterns highlighted in section 2. The rate of BAs at 25 among those from the most advantaged backgrounds is 75%, compared with 60% for those from the least advantaged. That the socio-economic attainment gap persists among these "high-achieving" students suggests the issue runs deeper than performance at school.

Wages. As the majority of the cohort members are in work at age 25, a focus of wave 8 is on their careers, occupations and other labour market outcomes. In particular they

[20]The survey is linked to an administrative education dataset, the National Pupil Database (NPD), so there is much more detailed information on the students' (academic) abilities potentially available. Unfortunately, I do not currently have access to this additional data as it must be accessed in the UK and only by researchers affiliated with a UK university.

Figure B1: Distribution of hourly wages at age 25, by degree attainment



Notes: The distributions are estimated (and plotted) using the `density` option in the `R` package `ggplot2` (Wickham, 2016), using the default setting of a Gaussian kernel density estimator. Analysis subsample ($N = 4,640$).

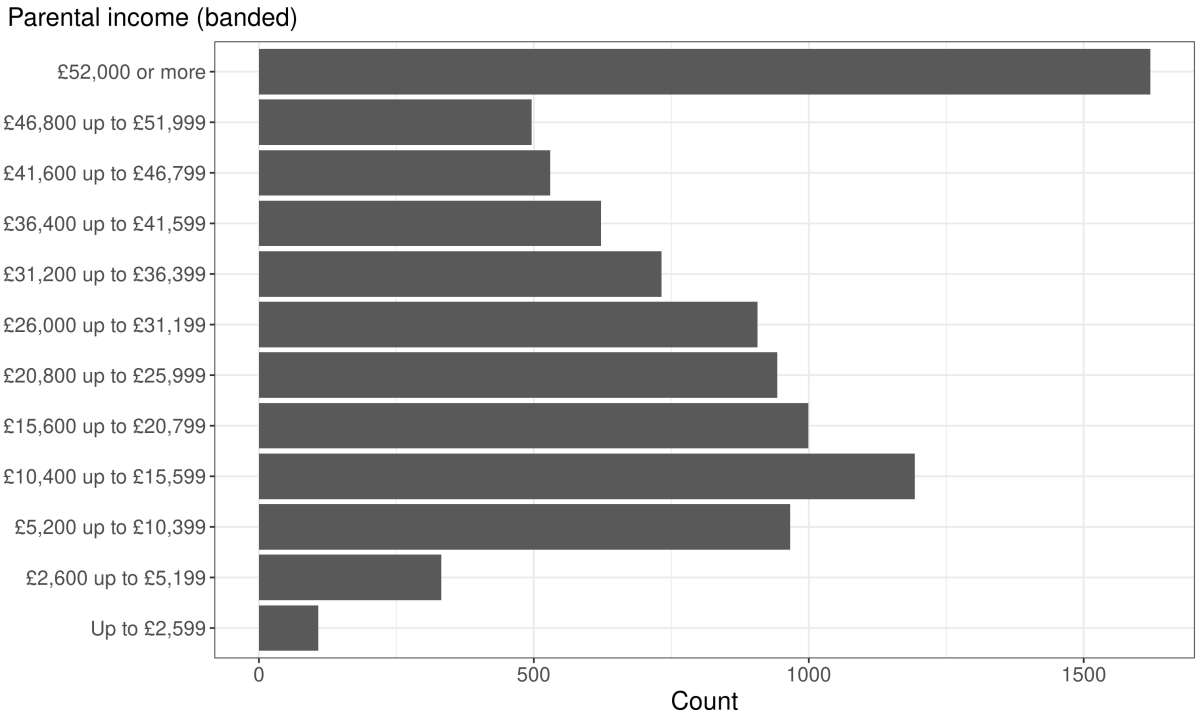
are asked to provide detailed information about their wages. Figure B1 shows the distribution of hourly wages in the sample, conditional on degree attainment. The conditional distributions look very similar, with the distribution corresponding to holders of an undergraduate (first) degree shifted slightly to the right. The mean and variance of these distributions are in table 1. However, such analysis does not reveal counterfactual wages: i.e. what graduates would earn had they not gone to university, and vice versa. For that we need the model and assumptions detailed in following sections.

B.3 Additional information on Next Steps

Next Steps started in 2004 when the members were in secondary school aged 13 or 14. They were then interviewed annually for the next six years, until aged 18 or 19 (waves 1–7). A further round of interviews (wave 8) was conducted in 2016 when the members were aged 25 or 26, and another is planned for 2021. For consistency with the BCS data, we will focus on the data collected at age 16 (or thereabouts, wave 4) and at 25 (wave 8).

Parental income. The LSYPE records information on member's family background in waves 1–7. Though data on parental income was collected in wave 4, it was recorded in 12 bins, with the top (and most populous) bin starting at £52,000 p.a. (see figure B2). More detail was collected in wave 1—over 30 bins, plus further information for some top-coded families—as well as continuous data on parents' salaries for some families (see figures B3

Figure B2: Parental income in the LSYPE (wave 4)



Source: LSYPE wave 4 (CLS, 2018).

and B4).

Undergraduate degree. Figure B6 shows the proportion of individuals in the LSYPE who hold a degree at 25, broken down by gender.

Table B2: Parameters selected by elastic net procedure

	$\lambda_{\min}$		λ_{1se}	
Variable name	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$
(Intercept)*	0.30	0.37	0.97	0.99
Ethnic group				
Mixed	.	0.05	.	.
Indian	0.20	0.24	.	.
Black Caribbean	−0.24	−0.26	.	.
Black African	0.22	0.28	.	.
Other	.	0.02	.	.

Table continues on next page ...

Table B2: (continued)

	$\lambda_{\min}$		λ_{1se}	
Variable name	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$
Number A−levels				
1	.	−0.07	.	.
2	−0.64	−0.65	−0.59	−0.55
4	0.44	0.43	0.11	0.09
5	0.78	0.76	0.29	0.24
6	0.48	0.50	.	.
Parental income				
Top 20%	0.04	0.06	.	.
Advantages				
Get better job	0.02	0.05	.	.
Well−paid job	.	0.03	.	.
Better opportunities	.	0.01	.	.
Need for career	−0.14	−0.17	.	.
Social life	−0.06	−0.09	.	.
Leave home	0.18	0.24	.	.
Personal development	0.05	0.09	.	.
Better life (general)	.	−0.04	.	.
Don't know	0.54	0.66	.	.
No answer	−0.15	−0.30	.	.
Disadvantages				
Get into debt	0.02	0.04	.	.
Depend on parents	.	0.01	.	.
Costs (general)	−0.05	−0.06	.	.
Takes long time	−0.06	−0.08	.	.
Waste of time	.	0.04	.	.
Degree = better pay				
Disagree	−0.04	−0.07	.	.
Owing money is wrong				
Agree	−0.11	−0.13	.	.
Borrowing money is normal				
Disagree	−0.12	−0.14	.	.
Debt difficult to get out of				
Disagree	0.09	0.11	.	.

Table continues on next page ...

Table B2: (continued)

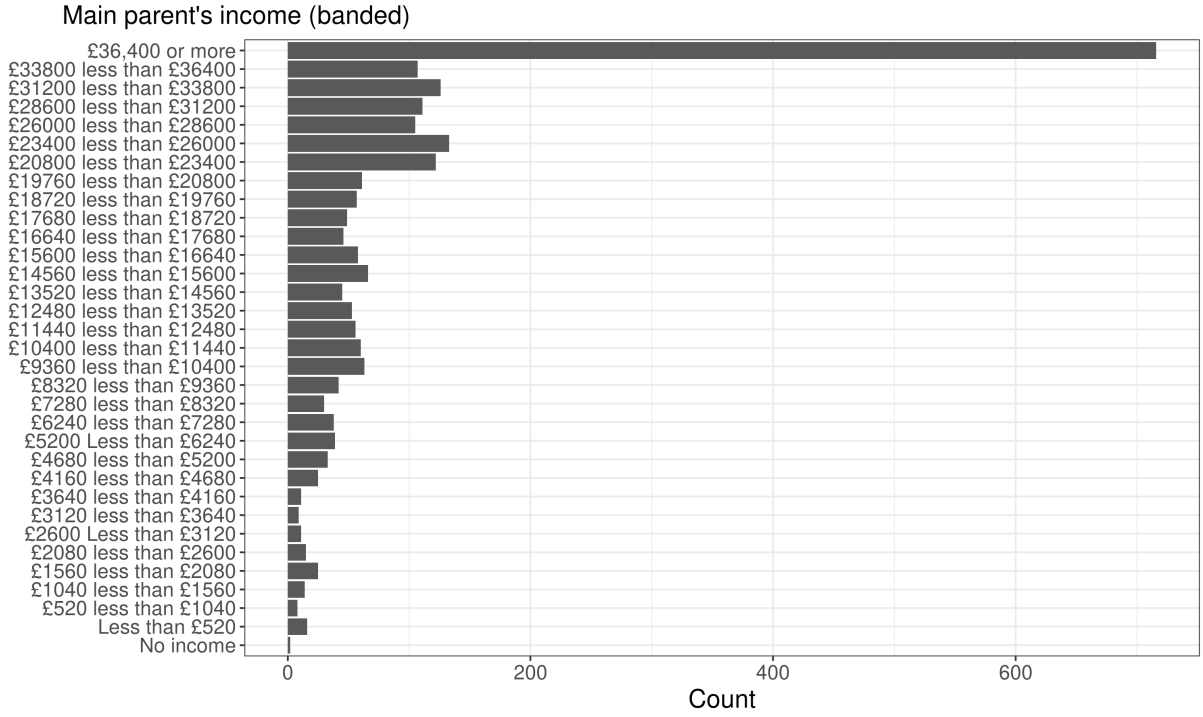
	$\lambda_{\min}$		λ_{1se}	
Variable name	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$
Student loan = cheap money				
Agree	0.04	0.05	.	.
Disagree	−0.05	−0.06	.	.
Need degree for job				
Agree	−0.02	−0.12	−0.31	−0.28
Disagree	0.47	0.39	.	.
Strongly disagree	0.62	0.52	0.14	0.13
Graduates get best jobs				
Disagree	−0.29	−0.30	−0.18	−0.17
Strongly disagree	−0.28	−0.34	−0.06	−0.06
Most friends going to uni				
Disagree	−0.29	−0.30	−0.20	−0.19
People like me don't go to uni				
Strongly disagree	0.31	0.31	0.22	0.19

*The intercept is not included in the selection / regularisation procedure.

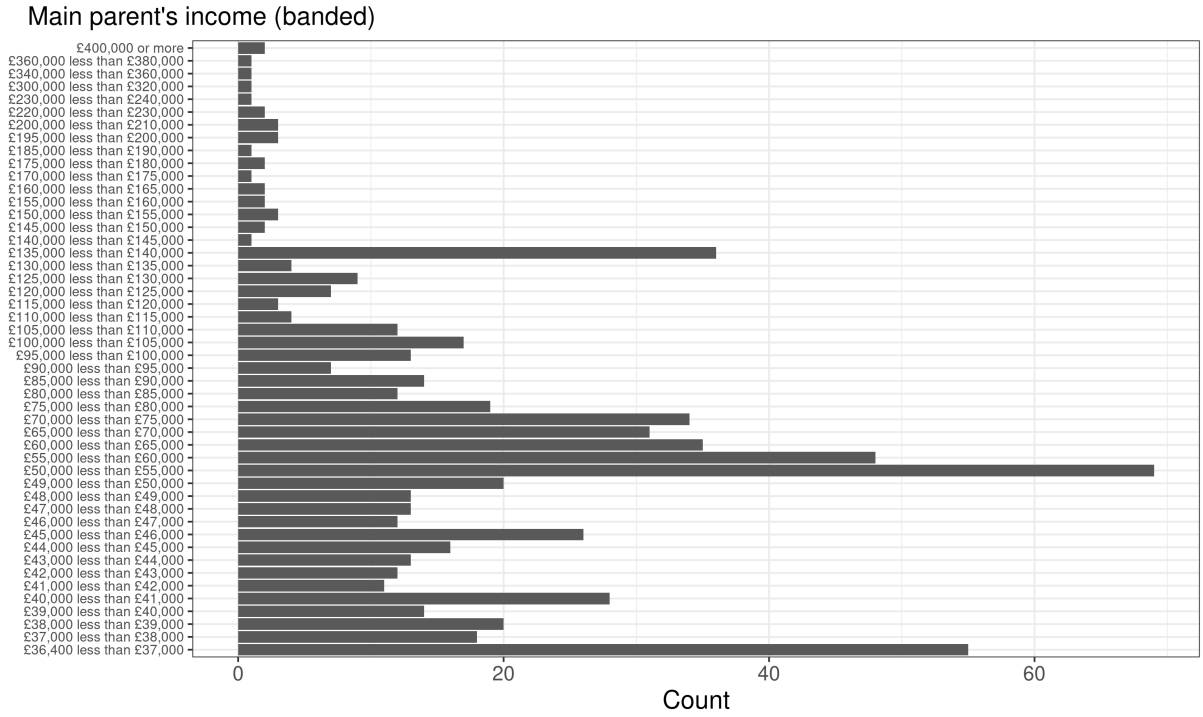
Notes: The elastic net procedure (and cross-validation to select λ values) was performed by the `cv.glmnet` function from the `R` package `glmnet` (Friedman et al., 2010). $\lambda_{\min}$ minimises the cross-validation error (binomial deviance here), while λ_{1se} selects the largest λ (corresponding to the fewest selected variables) with a cross-validation error within 1-standard error of the minimum.

Figure B3: Main parent's income in the LSYPE (wave 1)

(a) Banded



(b) Top-code ($> £36,400$) detail



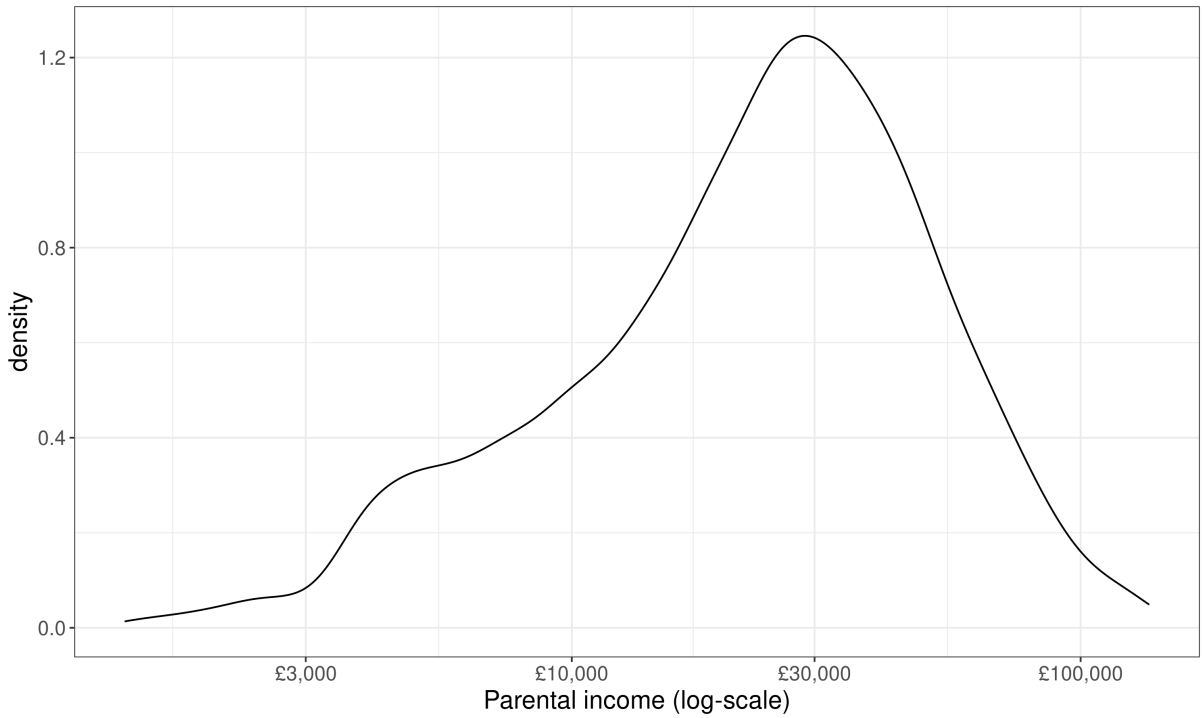
Source: LSYPE wave 1 (CLS, 2018).

Notes: The top panel (a) shows all recorded earnings for main parents in wave 1. Panel (b) shows a detailed breakdown of the top band from panel (a).

Table B1: The advantages (+) and disadvantages (−) of going to university

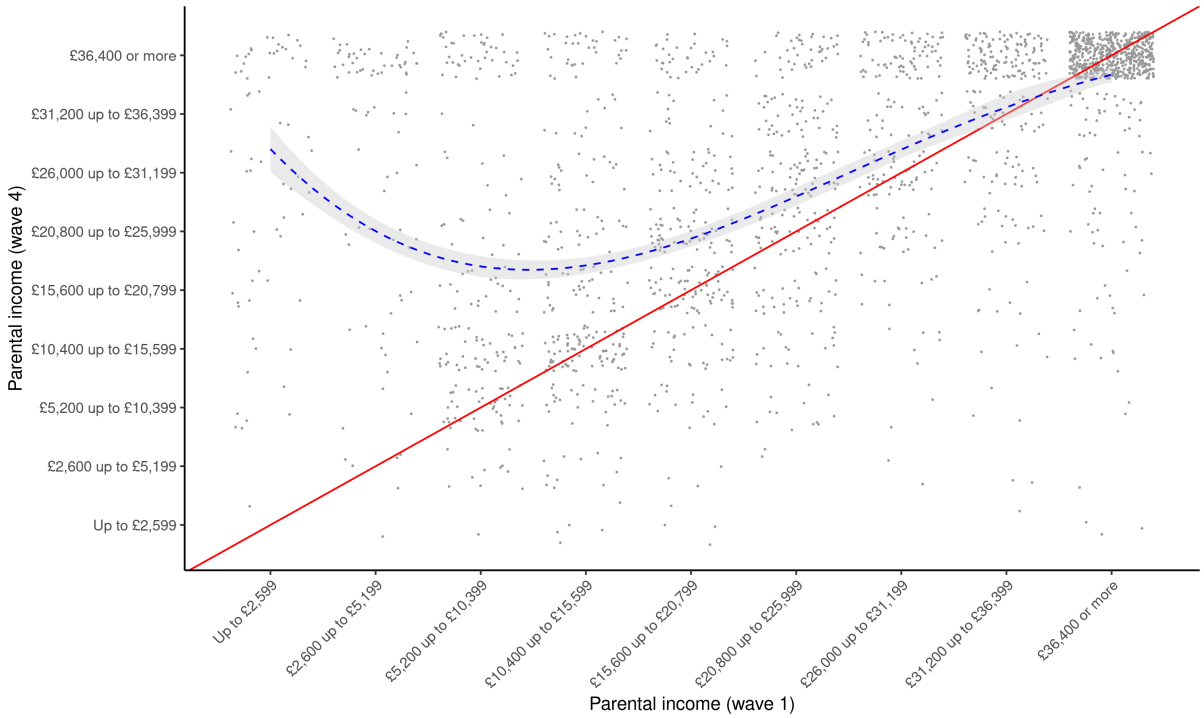
Response (harmonised)	+/−
Career	
Will lead to a good/better job (than would otherwise get)	+
Will lead to a well paid job	+
Gives someone better opportunities in life	+
Is essential for the career they want to go into	+
Shows that you have certain skills	+
To delay entering work/ more time to decide on a career	+
Not being able to start earning money/start work	−
No guarantee of a good job at the end	−
Don't need to go to university for the job someone may want	−
Get less work experience	−
Financial / debt	
Now	
It is expensive	−
Not becoming financially independent	−
Not being able to start earning money/start work	−
Costs (general/non specific)	−
Tuition fees/Accommodation costs/Living expenses	−
Future	
Will lead to a well paid job	−
Getting into debt/have to borrow money	−
Social life / environment	
The social life/ lifestyle / meeting new people / it's fun	+
To leave home/ get away from the area	+
Leaving home/family/friends	−
Stress	−
Education	
To carry on learning / I am good at / interested in my chosen subject	+
Get more qualifications/better/higher qualifications	+
The workload can be hard/ doubts about ability to finish course	−
Personal development	
Makes someone independent/ maturity / personal development	+
Gives you more confidence	+
People will respect me more	+
Leads to a better life/good life (general)	+
Prepare you for life/gain life skills	+
Time	
To delay entering work/ more time to decide on a career	+
Takes a long time	−
Waste of time (general/non-specific)	−

Figure B4: Parental income in the LSYPE (wave 1, density)



Source: LSYPE wave 1 (CLS, 2018).
Notes: This plot shows the density of (log-)annual earnings, calculated using the default kernel density estimator of the `geom_density()` function in the `ggplot2 R` package (Wickham, 2016).

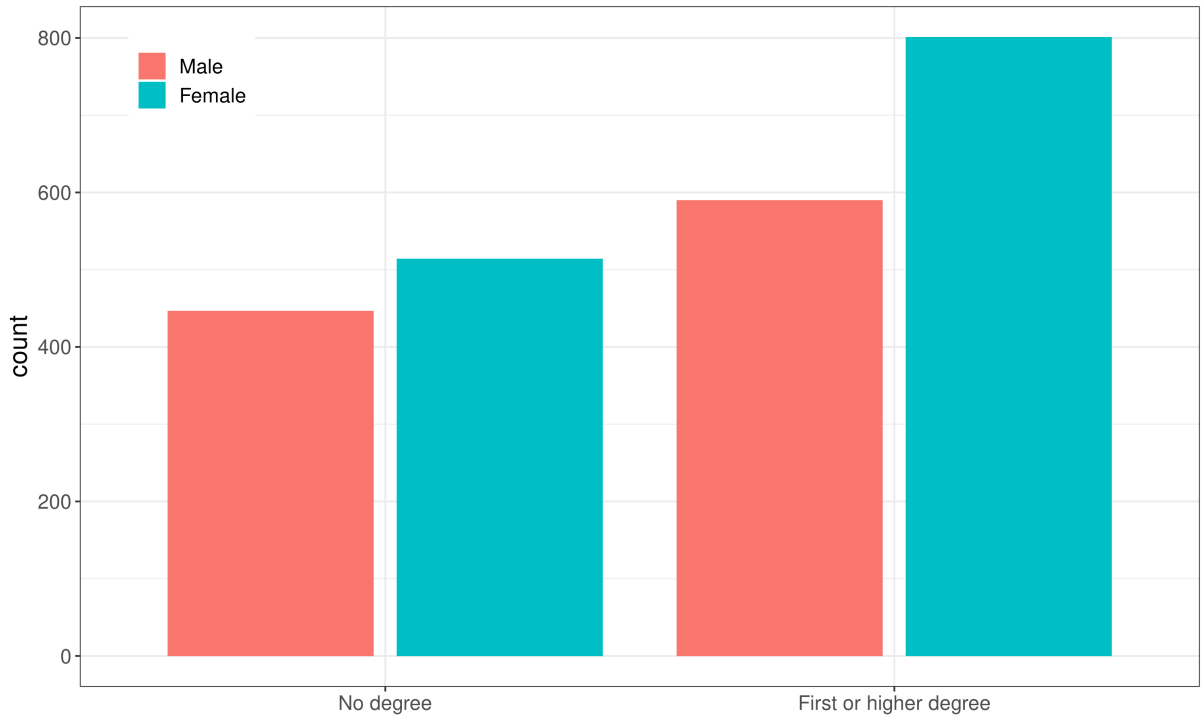
Figure B5: Comparing parental income in waves 1 and 4



Source: LSYPE (CLS, 2018).

Notes: This plot compares (banded) parental income in waves 1 and 4 (grey dots). To aid interpretation, a 2-knot spline (blue dashed line) is fitted using the `geom_smooth()` function in the `ggplot2 R` package (Wickham, 2016). The red line is the 45 degree line.

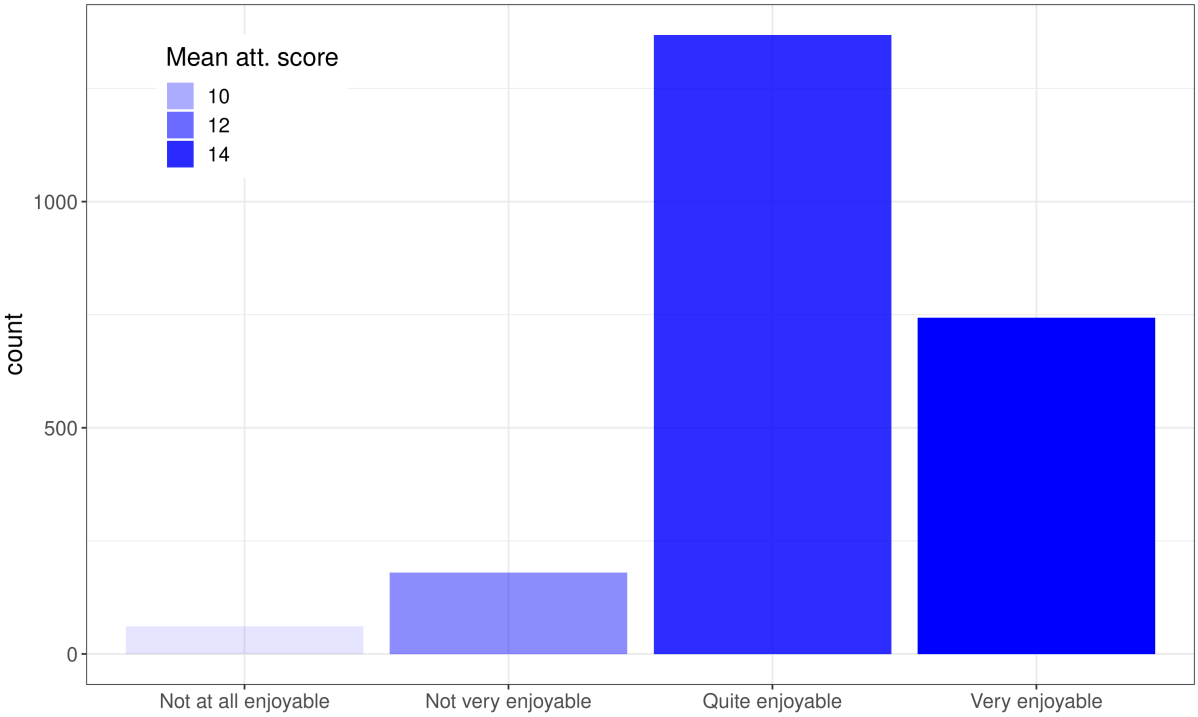
Figure B6: Undergraduate degree at 25 by gender (LSYPE)



Source: LSYPE wave 8 (CLS, 2018).

Figure B7: Students' enjoyment of year 11(a) and attitude towards school (b)

(a) How did you find year 11?



(b) Attitude towards school (1-20 scale)

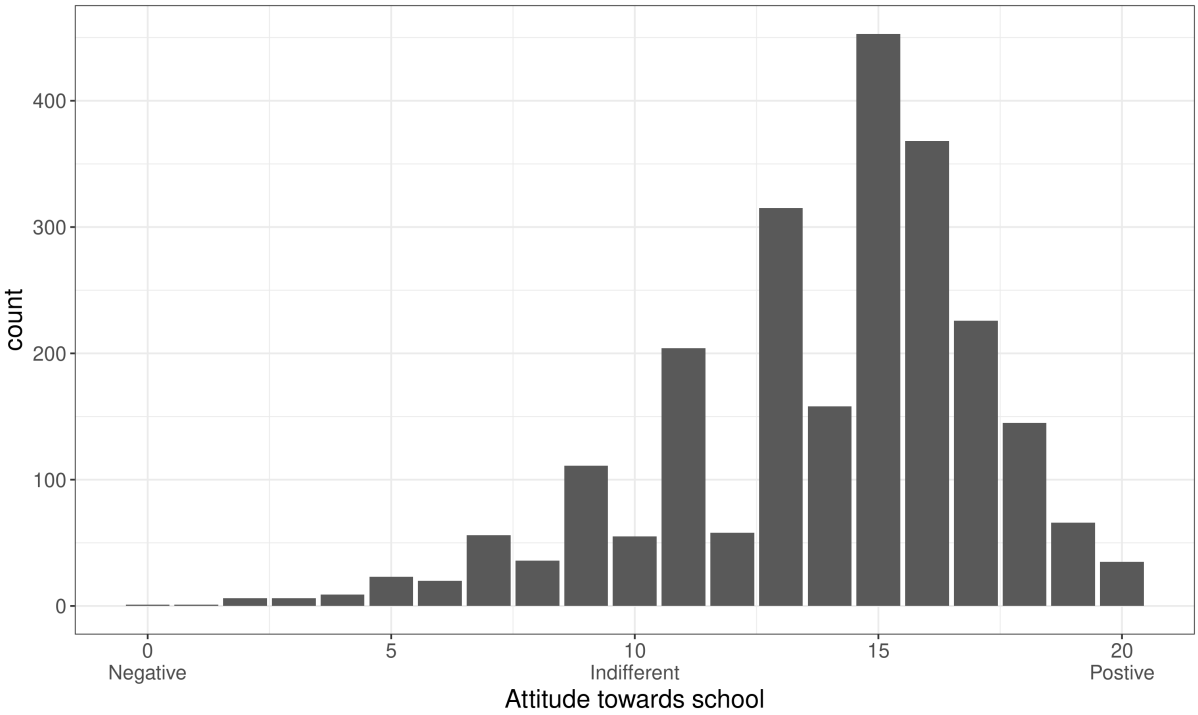
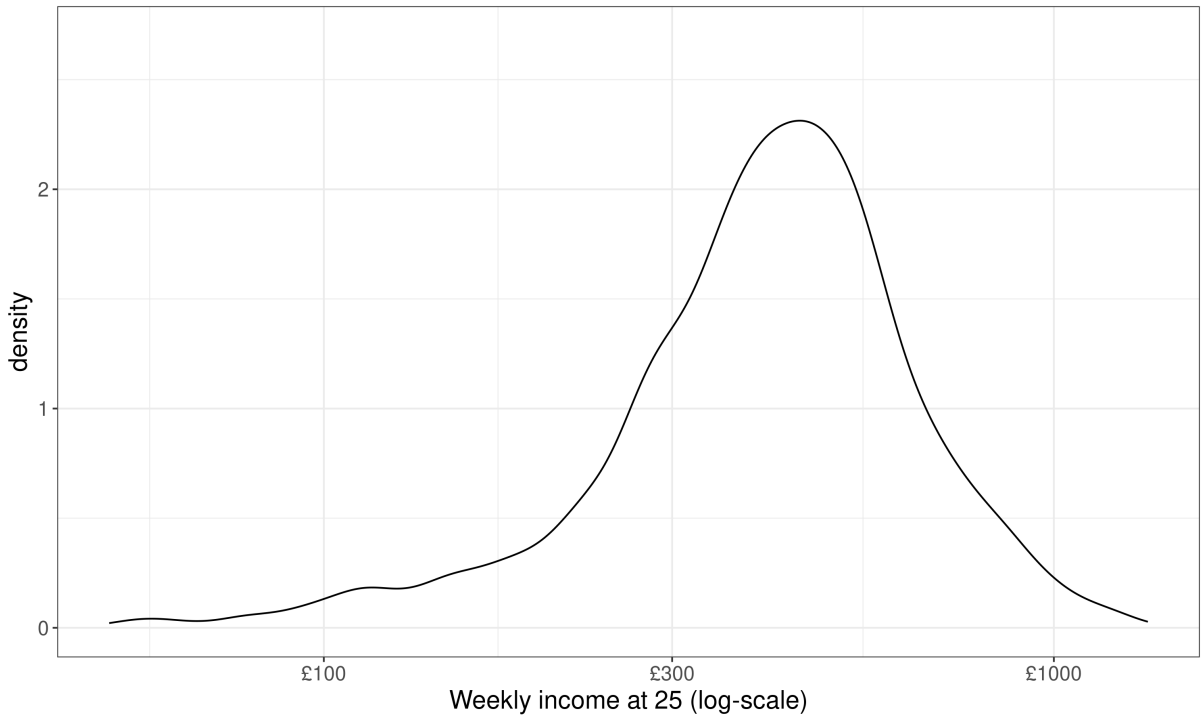


Figure B8: Distribution of weekly income at 25

(a) All



(b) Conditional on education

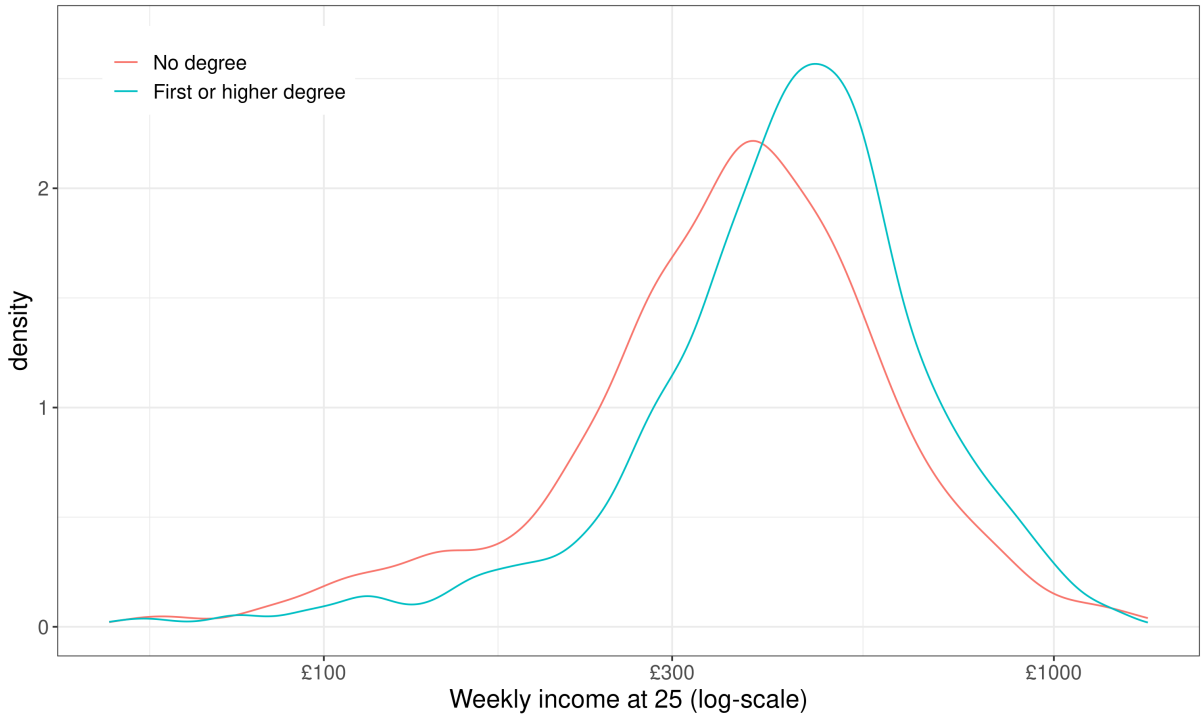


Figure B9: Responses classified as financial (all *disadvantages*)

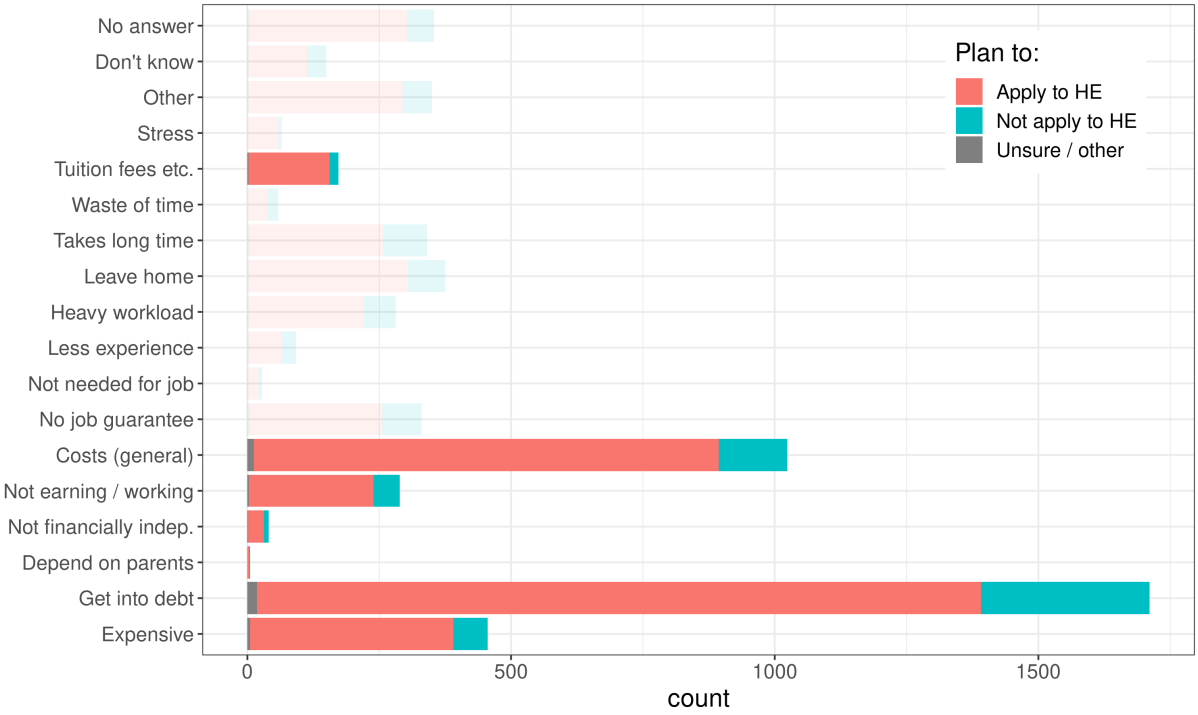
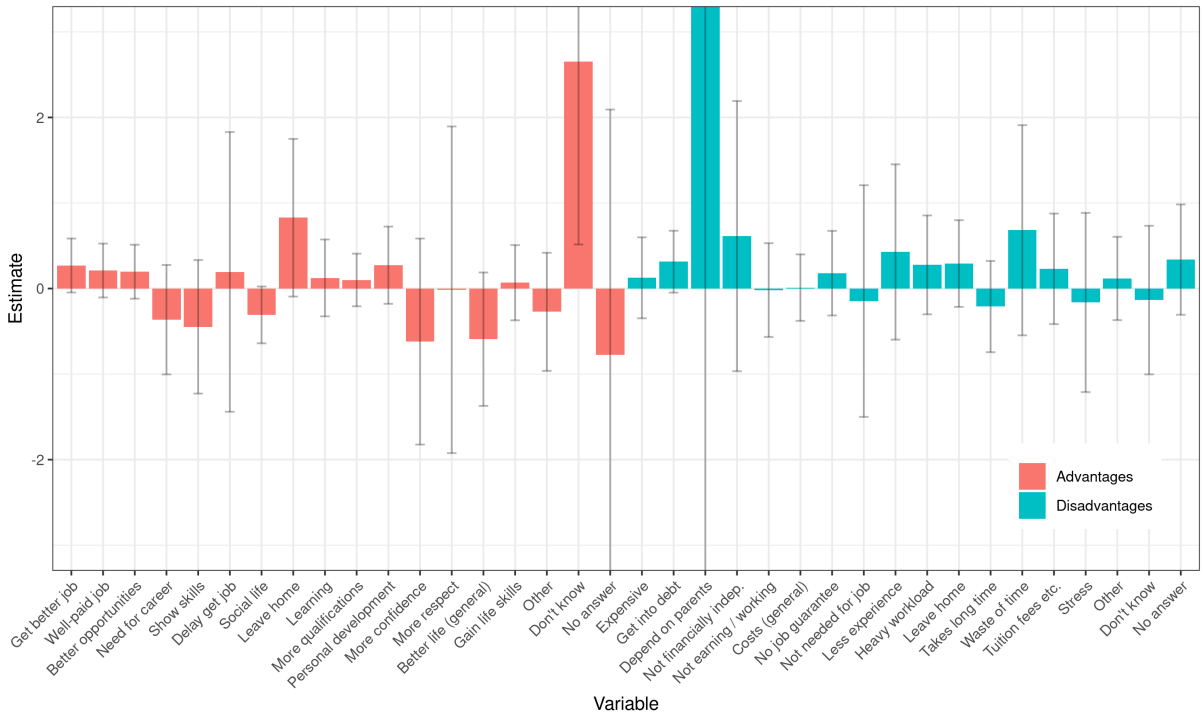
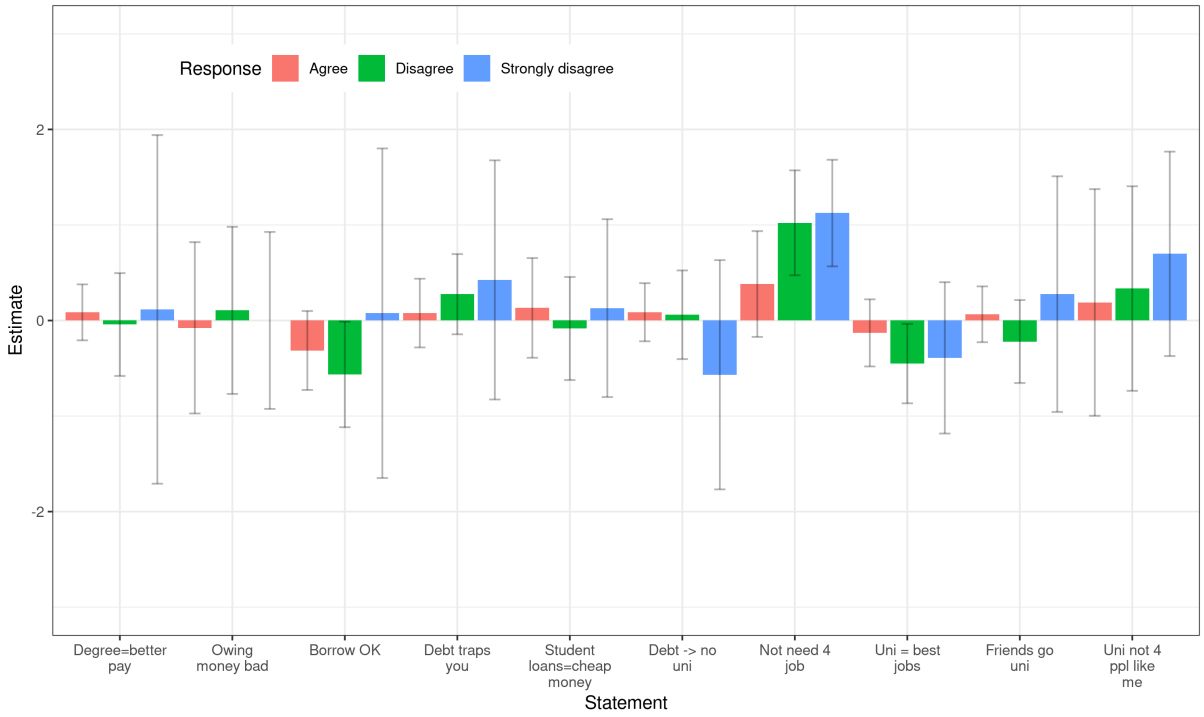


Figure B10: Parameter estimates for logit regression of university attendance

(a) Advantages and disadvantages of university



(b) Attitudes towards debt and higher education



Notes: For the advantages and disadvantages in panel (a) the reference category is not mentioning the corresponding advantage or disadvantage when surveyed. For the attitudes in panel (b), the reference category is to "strongly agree" with the corresponding statement. Statements 1–6 concern "debt", while 7–10 concern "higher education". Ethnicity, gender, number of A-levels, and parental income (in groups) were also included in the regression.

C More on identifying $\pi(k,z,s)$ and $f(\tilde{y}|k,s)$

We need the instrument z to produce variation in schooling for all types—i.e. the instrument must be valid for *all* types. Wage and test score distributions must differ across types; these are the sources of variation from which we identify types. Formally,

Assumption 1. $\pi(k,z,s) \neq 0, \forall s, \forall k, \forall z.$

Assumption 2. $\frac{\pi(k,1,d)}{\pi(k,0,d)} \neq \frac{\pi(k',1,d)}{\pi(k',0,d)}, \forall k \neq k', \forall d.$

Assumption 3. $\{g(\tau|k), k=1,...,K\}$ *and* $\{f(\tilde{y}|k,s), k=1,...,K\}$ *are linearly independent systems.*

Assume wages (and test score) are discrete,[21] taking N distinct values and form the matrix $P(z,s)$ which has test score in the rows, and wage at twenty-five in the columns. Each position (i,j) records the probability of observing test score i *and* wage at twenty-five j conditional on instrument z and schooling s. Fixing s and so dropping it from the notation, write

$$\underset{N\times N}{P(z)} = \begin{pmatrix} p(\tau_1,\tilde{y}_1,z) & p(\tau_1,\tilde{y}_2,z) & \cdots & p(\tau_1,\tilde{y}_N,z) \\ p(\tau_2,\tilde{y}_1,z) & p(\tau_2,\tilde{y}_2,z) & \cdots & p(\tau_2,\tilde{y}_N,z) \\ \vdots & \vdots & \ddots & \vdots \\ p(\tau_N,\tilde{y}_1,z) & p(\tau_N,\tilde{y}_2,z) & \cdots & p(\tau_N,\tilde{y}_N,z) \end{pmatrix} \tag{20}$$

We also stack the conditional income and wage probabilities in matrices.

$$\underset{N\times K}{G} = \begin{pmatrix} g(\tau_1|k=1) & g(\tau_1|k=2) & \cdots & g(\tau_1|k=K) \\ g(\tau_2|k=1) & g(\tau_2|k=2) & \cdots & g(\tau_2|k=K) \\ \vdots & \vdots & \ddots & \vdots \\ g(\tau_N|k=1) & g(\tau_N|k=2) & \cdots & g(\tau_N|k=K) \end{pmatrix}$$

$$\underset{N\times K}{F} = \begin{pmatrix} f(\tilde{y}_1|k=1) & f(\tilde{y}_1|k=2) & \cdots & f(\tilde{y}_1|k=K) \\ f(\tilde{y}_2|k=1) & f(\tilde{y}_2|k=2) & \cdots & f(\tilde{y}_2|k=K) \\ \vdots & \vdots & \ddots & \vdots \\ f(\tilde{y}_N|k=1) & f(\tilde{y}_N|k=2) & \cdots & f(\tilde{y}_N|k=K) \end{pmatrix}$$

Finally, call the diagonal $K \times K$ matrix of type probabilites, given z and for fixed s, D:

$$\underset{K\times K}{D(z)} = \begin{pmatrix} \pi(1|z) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \pi(K|z) \end{pmatrix} \tag{21}$$

[21] We can always discretise continuous wages and test scores using bins.

Then $P(z) = GD(z)F^\top$. Assumptions 1 and 3 ensure $P(z)$ has rank K, where K is the number of types. The key idea is that by manipulating the singular value decomposition (SVD) of this matrix, we can recover G and F, and $D(z)$ for each value of s. Reintroducing s into the notation, we obtain $F(s)$, $D(z,s)$, and $G(s)$. Note that $g(\tau|k,s=1) = g(\tau|k,s=0)$ as $g(\tau|k)$ is independent of s. Therefore we can use G to label types consistently across schooling levels—without doing this there is no reason that type 1 for $s=1$ should be the same type as type 1 for $s=0$.

D Estimating the extended model: further details

Recall the likelihood of obtaining i's observations, as a function of the parameters, $\boldsymbol{\Theta}$:

$$\ell(\boldsymbol{\Theta}; y_i, \tau_i, \boldsymbol{z}_i, X_{i,25}, s_i) = \sum_{k=1}^{K} \pi(k, \boldsymbol{z}_i, s_i) \cdot g(\tau_i|k) f(y_i|k, X_{i,25}, s_i). \tag{22}$$

The full sample likelihood is

$$\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{X}_{25}, \boldsymbol{s}) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi(k, \boldsymbol{z}_i, s_i) \cdot g(\tau_i|k) f(y_i|k, X_{i,25}, s_i), \tag{23}$$

and its logarithm

$$\begin{aligned}
\ln \mathcal{L}(\boldsymbol{\Theta}; W, \boldsymbol{Z}, \boldsymbol{Q}, \boldsymbol{s}) &= \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} \pi(k, \boldsymbol{z}_i, s_i) \cdot g(\tau_i|k) f(y_i|k, X_{i,25}, s_i) \right) \\
&= \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_k \cdot \underbrace{\pi(\boldsymbol{z}_i, s_i|k) \cdot g(\tau_i|k) f(y_i|k, X_{i,25}, s_i)}_{\ell(\boldsymbol{\Theta}; y_i, \tau_i, \boldsymbol{z}_i, X_{i,25}, s_i, k)} \right)
\end{aligned}$$

where $p_k = \pi(k)$ the unconditional probability of an individual being type k. The MLE estimator of the parameters, Θ, satisfies

$$\hat{\Theta} \equiv \arg\max_{\Theta} \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_k \ell(\boldsymbol{\Theta}; y_i, \tau_i, \boldsymbol{z}_i, X_{i,25}, s_i, k) \right) \tag{24}$$

The sum inside the logarithm prohibits sequential estimation of the parameters in Θ.

Arcidiacono and Jones (2003) show the same $\hat{\Theta}$ satisfies

$$\hat{\Theta} \equiv \sum_{i=1}^{N} \sum_{k=1}^{K} \Pr(k|y_i, \boldsymbol{z}_i, X_{i,25}, s_i; \hat{\Theta}, \hat{p}) \ln \ell(\Theta; y_i, \tau, \boldsymbol{z}_i, X_{i,25}, s_i, k) \tag{25}$$

where

$$p_i(k|\Theta) \equiv \Pr(k|y_i, \boldsymbol{z}_i, X_{i,25}, s_i; \Theta, p) = \frac{p_k \ell_i(\Theta; y_i, \boldsymbol{z}_i, X_{i,25}, s_i, k)}{\sum_{k=1}^{K} p_k \ell_i(\Theta; y_i, \boldsymbol{z}_i, X_{i,25}, s_i, k)} \tag{26}$$

and

$$\hat{p}_k = \frac{1}{N} \sum_{i=1}^{N} \Pr(k|y_i, \boldsymbol{z}_i, X_{i,25}, s_i; \hat{\Theta}, \hat{p}). \tag{27}$$

Crucially, the right-hand side of (25) *lends itself to sequential estimation.*

D.1 The EM algorithm

The EM algorithm takes its name from the two steps over which the algorithm iterates: an expectation (E) step, and a maximisation (M) step. Below we detail the E- and M-steps for the $m+1$-th iteration of the algorithm adapted to estimate the cross-cohort model. Subscripts on parameters denote the iteration from which they were estimated, so $\beta^{(m)}$ is the estimate of β obtained in the m-th iteration. Recall we are trying to find Θ which solves

$$\hat{\Theta} \equiv \arg\max_{\Theta} \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_k \ell(\Theta; y_i, \tau, \boldsymbol{z}_i, X_{i,25}, s_i, k) \right), \tag{28}$$

where

$$\Theta = \left(\{\alpha_{k1}, \alpha_{k0}, \mu_k\}_{k=1}^{K}, \beta_1, \beta_0, \sigma_\varepsilon, \sigma_\nu \right). \tag{29}$$

To obtain the $m+1$-th estimates given the m-th estimates, proceed as follows.

E-step. In the E-step we update the posterior type probabilities, $p_i(k|\Theta)$:

$$p_i\left(k|\hat{\Theta}^{(m)}\right) \equiv \frac{\hat{p}_k^{(m)} \ell\left(\hat{\Theta}^{(m)}; y_i, \tau_i, \boldsymbol{z}_i, X_{i,25}, s_i, k\right)}{\sum_{k=1}^{K} \hat{p}_k^{(m)} \ell\left(\hat{\Theta}^{(m)}; y_i, \tau_i, \boldsymbol{z}_i, X_{i,25}, s_i, k\right)}. \tag{30}$$

M-step. While in the M-step we update the components of Θ:

- Update μ_k, σ_ν.
    1. Update μ_k as the weighted mean test score, using posterior probabilities as weights (for each type)

$$\mu_k^{(m+1)} \equiv \frac{\sum_i p_i\left(k|\hat{\Theta}^{(m)}\right) \tau_i}{\sum_i p_i\left(k|\hat{\Theta}^{(m)}\right)} \tag{31}$$

    2. Then σ_ν is updated as the weighted root-mean-square error, using posteriors

as weights (over all types)

$$\sigma_{\nu}^{(m+1)} \equiv \sqrt{\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N} p_i\left(k|\hat{\Theta}^{(m)}\right)(\tau_i - \mu_k)^2} \tag{32}$$

- Update β_s, σ_{ε}, and α_{ks}.

  1. Use weighted least squares (LS) of $\ln y - \alpha_{ks}$ on X_{25}, with weights $p_i(k|\hat{\Theta}^{(m)})$ to update β_s for each value of s

  $$\hat{\beta}_{js}^{(m+1)} \equiv \frac{\sum_{i\in I(s)}\sum_{k=1}^{K} p_i\left(k|\hat{\Theta}^{(m)}\right) X_{ij,25}\left(\ln w_i - \alpha_{ks}^{(m)}\right)}{\sum_{i\in I(s)}\sum_{k=1}^{K} p_i\left(k|\hat{\Theta}^{(m)}\right) X_{ij,25}^2} \tag{33}$$

  where $\beta_s = (\beta_{1s}, ..., \beta_{Js})$, i.e. j indexes β_s and X_{25}, and $I(s) = \{i : S_i = s\}$.

  2. Then use the updated β_s to update α_{ks} for given k, s:

  $$\hat{\alpha}_{ks}^{(m+1)} \equiv \frac{\sum_{i\in I(s)} p_i\left(k|\hat{\Theta}^{(m)}\right)\left(\ln w_i - X_{i,25}'\beta_s^{(m+1)}\right)}{\sum_{i\in I(s)} p_i\left(k|\hat{\Theta}^{(m)}\right)} \tag{34}$$

  3. And use the updated α_{ks} and β_s to update $\sigma_{\varepsilon}(k)$:

  $$\hat{\sigma}_{\varepsilon}^{(m+1)}(k) \equiv \sqrt{\frac{\sum_{i\in I(s)} p_i\left(k|\hat{\Theta}^{(m)}\right)\left(\ln w_i - X_{i,25}'\beta_s^{(m+1)} - \alpha_{ks}^{(m+1)}\right)^2}{\sum_{i\in I(s)} p_i\left(k|\hat{\Theta}^{(m)}\right)}} \tag{35}$$

- Finally, we sum posterior probabilities by k, $\boldsymbol{z}$, and s to obtain $\pi(k, \boldsymbol{z}, s)$,

$$\hat{\pi}^{(m+1)}(k, \boldsymbol{z}, s) \equiv \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I(k,\boldsymbol{z},s)} p_i(k|\hat{\Theta}^{(m)}), \tag{36}$$

  where $I(k, \boldsymbol{z}) = \{i : \boldsymbol{z}_i = \boldsymbol{z}, s_i = s\}$.

Iterations stop when the algorithm converges, i.e. when the increase in likelihood between iterations is below a threshold. Formally, stop at iteration m if

$$\mathcal{L}(\boldsymbol{\Theta}^{(m)}; \boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{Z}, \boldsymbol{X}_{25}, \boldsymbol{s}) - \mathcal{L}(\boldsymbol{\Theta}^{(m-1)}; \boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{Z}, \boldsymbol{X}_{25}, \boldsymbol{s}) < \delta, \tag{37}$$

for some $\delta > 0$ chosen by the econometrician.