The influence of socioeconomic status on changes to young people's expectations of applying to university*

Jake Anders†

Conference submission:
12 December 2014

Abstract

A much larger proportion of English 14 year olds expect to apply to university than ultimately make an application by age 21, but the proportion expecting to apply falls from age 14 onwards. In order to assess the role of socioeconomic status in explaining changes in expectations, I apply duration modelling techniques to data from the Longitudinal Study of Young People in England, analysing transitions in young people's expectations both from being 'likely to apply' to being 'unlikely to apply' and vice versa. I find that young people's socioeconomic background has a significant association with changes in expectations, even once I control for prior academic attainment and other potentially confounding factors. This suggests more could usefully be done to maintain the educational expectations of academically able young people from less advantaged families. Furthermore, young people's backgrounds affect their responsiveness to new evidence on academic attainment at age 16, contributing to the socioeconomic gradient in expectations.

Keywords: Educational expectations, socioeconomic status, Higher Education, Duration modelling.
JEL Classification Numbers: I24.

†This research forms part of the Nuffield Foundation project "Higher Education Funding and Access: exploring common beliefs" (grant number EDU/39084). The project is based at the Institute of Education and the Institute for Fiscal Studies. It is directed by Claire Crawford, Lorraine Dearden, John Micklewright and Anna Vignoles. Thanks to my supervisors Lorraine Dearden and John Micklewright for their patient support. I gratefully acknowledge my PhD scholarship funded by the IoE as part of the ESRC NCRM ADMIN node.
†National Institute of Economic and Social Research/UCL Institute of Education, University College London, jake@jakeanders.uk

1 Introduction

In previous work, I found a large socio-economic gradient in university application in England. Much of this inequality can be explained by differences in academic achievement that emerge long before the point at which young people apply to university (see also Chowdry et al., 2013). However, even conditioning on these earlier academic outcomes and other potential confounding factors, a socio-economic gradient in whether or not individuals make an application to university remains (Anders, 2012a). This is despite the fact that a larger proportion of English 14-year-olds from disadvantaged backgrounds expect to apply to university than the overall proportion who have ultimately done so by age 21 (Anders and Micklewright, 2013, pp.42-43).

This raises the question of when and why young people from less advantaged families change their minds about making an application to university. Are their changes in expectations explicable by other factors, such as academic attainment, or does socio-economic status continue to have an influence? Given the previous evidence that much of the socio-economic gap in university attendance opens at or before the point of application, a better understanding of the dynamics of whether or not individuals expect to apply is of significant importance to the formulation of policy on reducing the socio-economic gradient in access to Higher Education.

Rather than following previous authors in using expectations data as an explanatory factor for later outcomes, in this paper I take a step back, addressing the issue directly by analysing the influence of socio-economic status on the large number of changes in young people's expectations of applying to university between ages 14 and 17, just before young people start making applications to university. Using rich panel data from the Longitudinal Study of Young People in England (LSYPE), I take the novel approach of using duration modelling to analyse the dynamics of young people's expectations.

The research question and data used lend themselves naturally to this approach. Duration modelling allows the flexibility to make use of all available information on the timing of events (including the possibility of multiple transitions back and forth between reporting 'likely' and 'unlikely' by an individual), it can take account of changes in young people's circumstances during the period under consideration, and allows for more flexible handling of some missing outcomes data. The technique also allows separate analysis of both transitions from being 'likely to apply' to being 'unlikely to apply' and vice versa. This is important, since the factors which cause young people to raise their expectations and start thinking that they are likely to apply to university may be quite different from the causes of movement in the other direction. Despite this, duration modelling is not regularly used in such settings and, to my knowledge, has not been used before to model changes in young people's educational expectations over time.

This paper makes an important contribution to the literature on access to Higher Education. Using the longitudinal nature of the data, I provide non-parametric estimates of changes in young people's expectations between the ages of 14 and 17, quantifying the extent of changes in expectations during this period. Making minimal assumptions, I also use this technique to examine whether young people from less advantaged backgrounds are more likely to stop, and less likely to start, thinking they are likely to apply to university than their more advantaged peers. Furthermore, taking advantage of the rich survey data and retaining the flexibility of duration modelling, I provide estimates of the continued influence of socio-economic status, after controlling for potentially confounding factors including prior academic attainment and demographic characteristics. Finally, I explore the interplay between SES and new information on academic attainment at age 16.

The paper proceeds as follows. Section 2 reviews the literature on the socio-economic patterning of educational expectations and lays out a modelling strategy for identifying the influence of socio-economic status on changes in expectations. Section 3 describes the dataset and measures used in this paper. Section 4 introduces duration modelling as applicable to these data and sets out the benefits of using it to analyse changes in expectations. Non-parametric duration modelling methods are applied in Section 5 to explore how young people's expectations change during their teenage years and how this is associated with socio-economic status. This initial analysis is extended through use of multiple regression models, introduced in Section 6 and with the results of this analysis reported in Section 7. Finally, Section 8 concludes.

2 Background and identification strategy

This paper, rather than attempting to identify the effect of young people's expectations on university attendance, takes a step back. It explores the role of socio-economic status (SES) in determining the paths of young people's expectations in the first place. The importance of young people's expectations, particularly in explaining the SES gradient in academic attainment, has increasingly attracted academic interest over the past few years. This has been accompanied by policy makers emphasising the need to 'raise aspirations', particularly among high attaining, but low SES, young people.[1] Such policies, in the UK, have included the now-defunct 'Aimhigher' programme and requirements for outreach work by universities charging more than £6,000 in tuition fees in their Access Agreements with the Office For Fair Access (OFFA).

[1]A DfE-funded study reflecting this concern found that most schools it surveyed indicated that "encouraging their students to apply to higher education [...] was one of their highest priorities" (Thornton et al., 2014, p.146).

It is important to distinguish upfront between young people's expectations and their aspirations. Jerrim (2011, p.6-7) summarises the difference between the two as being that expectations "implies a realistic assessment of future outcomes, while [aspirations] reflects children's hopes and dreams". For this paper's application, young people might hope to apply to university (an aspiration), without expecting that they will be in a realistic position to do so. Although much of the policy discourse focuses on 'raising aspirations' rather than 'raising expectations', expectations seem more likely to be informative for the purposes of this paper, but understanding both aspirations and expectations pose many of the same challenges.

Regardless of the interest of policymakers, studying expectations is not worthwhile if they are just an individual's whim. However, Morgan (1998) argues that "educational expectations are not 'flights of fancy' or 'vague preferences' [but rather,] because they can be explained by a reasonable theory of rational behavior, should be considered rational" (Morgan, 1998, p.157) and hence, presumably, informative. Certainly, previous work has shown a correlation between educational expectations and later outcomes. Chowdry et al. (2011) find a correlation between young people thinking it likely that they will apply to university and academic performance at age 16, even after controlling for long-run family background factors and prior attainment. Elsewhere in the world, analysis of the Longitudinal Survey of Australian Youth estimates that the "correlation between intention and entry to higher education is moderately strong (r = 0.59)" (Khoo and Ainley, 2005, p.v). Similarly, in the US, Reynolds and Pemberton (2001) report that while 29% of those who expect to complete a college degree when asked in 1979 (age 15-16) had done so by 1994 (aged 30-31), under 3% of those who did not expect to complete a college degree had done so (Reynolds and Pemberton, 2001, p.723).

Using data from the Programme of International Student Achievement (PISA) survey, Jerrim (2011) examined the socio-economic patterning of young people's expectations of completing Higher Education. He finds that that there are large differences between advantaged and disadvantaged children's expectations in most countries throughout the developed world. He finds that England is no exception to this pattern, with only a handful of OECD countries having significant differences (on either side) in the strength of the relationship. By contrast, the correlation between socio-economic advantage and expectations is significantly weaker in the US than most other OECD countries, including England.

Why do these associations between expectations and outcomes exist? One potential explanation is that young people who grow up in more deprived households "may expect less of themselves and may not fully develop their academic potential because they see little hope of ever being able to complete college or using their schooling in any effective way" (Cameron and Heckman, 1999, p.86).

However, others, such as Gorard (2012), are highly critical of the jump from these plausible explanations and observed correlations between attitudes and academic outcomes to seeing the relationship as playing a truly causal role. Gorard argues that formulating policy on this basis, when evidence of causation is so weak, is misguided because of the opportunity costs and potential negative side effects of policies aimed at raising aspirations and expectations.

Given this paper's focus on the influence of SES on the pathways of young people's expectations, expectations data are used as an outcome variable. Doing so means taking a step back from its use as an explanatory variable, as was the case in the studies above. The focus on expectations as an outcome variable means that there is no need to take a view on whether or not expectations have a causal impact on academic attainment and progression. Instead, it is enough to be convinced that young people's expectations are at least symptomatic of the underlying social processes leading from SES, prior attainment, and other background characteristics to the ultimate decision as to whether or not to apply to university.

This paper contributes to a literature on the formation and correlates of young people's educational expectations and aspirations. Previous work has considered similar issues in differing contexts or applying differing methods. However, this is the first analysis to consider a dynamic relationship between SES and young people's expectations. Rampino and Taylor (2013) analyse young people's educational aspirations using data from the British Household Panel Study (BHPS), focusing in particular on differences by gender, using responses to questions such as "Would you like to go on to do further full-time education at a college or University after you finish school?".[2] They do not consider changes in aspirations, but do take advantage of the panel nature of the data, estimating probit models with individual-level random effects. Fumagalli (2012) also estimates binary choice models of young people's expectations of getting a place at university (with adjustment for selection effects in who is asked the question of interest) using the same dataset as that which I use. Perhaps the paper closest in aims to this paper is Kao and Tienda (1998): using data from the US, they estimate logistic regression models of the association between young people's background characteristics and changes in educational aspirations (including an aspirations variable lagged by one time period as a covariate).

These previous studies have all found a role for socio-economic status. Kao and Tienda find that socio-economic background "exerts a strong influence on educational aspirations and is vital to their maintenance through the high school years" (Kao and Tienda, 1998, p.370). Rampino and Taylor

[2]The BHPS lacks data on young people's prior academic attainment, which is available in the dataset used in this paper, and which would be strongly expected to be relevant to educational expectations.

report that “the educational aspirations of boys are more positively affected by parental education than those of girls” (Rampino and Taylor, 2013, p.34), also noting that the effect of parental attitudes varies by gender in the same way. Fumagalli finds that young people from families with higher parental education are more responsive to new information about their academic attainment in updating their expectations of both applying to university and ultimately getting a place. In addition, she finds that, contrary to popular belief, “young people from free school meal eligible families have more positive expectations [of being accepted to university, conditional on having applied], even when grades are controlled for” (Fumagalli, 2012, p.41-42).

This paper builds on the previous literature in two important respects. First, through use of duration modelling, this paper analyses the dynamic relationship between SES and young people’s expectations in a flexible way. Importantly, it allows for different relationships between characteristics of interest and whether young people make a transition depending on direction of the transition (i.e., ‘likely to unlikely’ or ‘unlikely to likely’). Second, both Kao and Tienda and Rampino and Taylor focus on aspirations rather than expectations, while Fumagalli analyses formation of young people’s expectations of being admitted to university, conditional on having made an application.[3] Here, the focus is on expectations of applying to university, which is distinct from any of these.

To analyse the influence of SES on the likelihood of changes in young people’s expectations, one must first have some idea of the relationship between the two. Drawing on others’ findings about the determinants of expectations (for example Kao and Tienda, 1998; Fumagalli, 2012; Anders and Micklewright, 2013; Rampino and Taylor, 2013) I treat the probability of transition as a function of SES and various other characteristics:

$$Pr(\Delta\text{Expectations}) = f(\text{SES}, X) \tag{1}$$

where X is a vector of characteristics including young people’s age, academic ability, demographic characteristics, school characteristics, traumatic experiences, and local labour market conditions.

The strategy is to isolate the role of SES by controlling for elements of X. However, there are several challenges to achieving this. Several of these are discussed in Section 3.4 below, where the measurement of these variables in the dataset is considered. Most fundamentally, one cannot be sure that other unobserved or unobservable elements do not also appear in the function. In the absence of exogenous variation in SES (which is conceptually, let alone practically, challenging) one cannot

[3] As the question on likelihood of admission, conditional on application, is only asked to individuals who indicate that they are more than ‘not at all likely’ to apply, Fumagalli does estimate models of likelihood of applying (focusing on the probability of being at least ‘not very likely’ to apply) to deal with this selection problem.

be certain that this problem has been dealt with. However, an alternative strategy, making use of random effects (modelled either as having a normal distribution or a discrete mixing distribution), to help deal with unobserved heterogeneity is discussed and applied in Appendix C. The results obtained when I apply this method do not substantively alter the findings from this analysis in this chapter, giving me some confidence in the qualitative story from my estimates.

3 Data

The Longitudinal Study of Young People in England (LSYPE) is a major panel survey, funded to age 20 by the UK Department of Education. The LSYPE tracks the experiences of one cohort of young people over seven years (with one interview per year), from approximately age 14 (in 2004) to age 20 (in 2010),[4] including interviews with the young people themselves (throughout) and their parents (up to age 17). It collected a wide variety of data on participants, including details on their socio-economic background, educational attainment, and educational expectations. Only aspects of the LSYPE relevant to the research questions of this paper are discussed here; more in depth description of the LSYPE is available in Anders (2012b).

As with any longitudinal survey, the LSYPE suffers from attrition. One of the advantages of duration modelling is the option of treating missing outcome data as 'censored' (discussed further in Section 4). This is preferable to having to drop respondents that attrit from from the analysis, which would mean being restricted to a complete case sample of 8,029.[5] Individuals who are not present in both Waves 1 and 2 are excluded, to ensure that at least one potential transition is observed for all individuals included the analysis. The number of participants at Wave 2 is 13,447 out of the 15,770 who initially responded at Wave 1 (i.e. an 85% response rate). However, missing data for key variables reduce the sample size in the analyses to those reported in the results tables. I weight the data for my analysis using the LSYPE-provided attrition and non-response weights for Wave 2.

This section discusses four main aspects of the data. First, the measurement of the outcome variable (young people's expectations of applying to university), including specifics of measurement in this dataset and more general challenges posed by use of expectations data as an outcome in duration modelling. Second, the sequences of expectations observed in the data. Third, the measurement of the main explanatory variable of interest (young people's SES), including construction of an index of SES from various indicators. Finally, the measurement of other characteristics that may confound the

[4]Further waves following the young people as they enter the labour market are now planned, funded by the Economic and Social Research Council. For more information visit `http://www.cls.ioe.ac.uk/lsype`.

[5]This complete case sample is used (applying appropriate attrition weights) in Figure 1 and as a robustness check, reported in Appendix B.

relationship between SES and changes in expectations.

3.1 Measurement of expectations

The LSYPE begins recording young people's expectations of applying to university from approximately age 14. Conveniently, given that this is the earliest point in the data, previous psychological and sociological literature has argued that this is also the age at which young people "relinquish their most preferred [occupational] choices and settle for more acceptable, available, choices" (Gutman and Akerman, 2008, p.5). Similarly, Gottfredson (2002, p.98-101) argues that by the age of 14, young people have completed 'circumscription' of their aspirations, whereby they rule out unacceptable career aspirations, and begin 'compromise' by "adjusting their aspirations to accommodate an external reality" (Gottfredson, 2002, p.100). It follows that age 14 is a natural point from which to analyse young people's expectations in a meaningful way; as such, I treat young people's periods of reporting their expectations as starting at this point at the earliest.

The LSYPE measures young people's expectations of applying to university through a single question repeated in most of the waves of the survey. Young people are asked "How likely do you think it is that you will apply to university?" and are asked to choose from the options 'very likely', 'fairly likely', 'not very likely',[6] and 'not at all likely'.

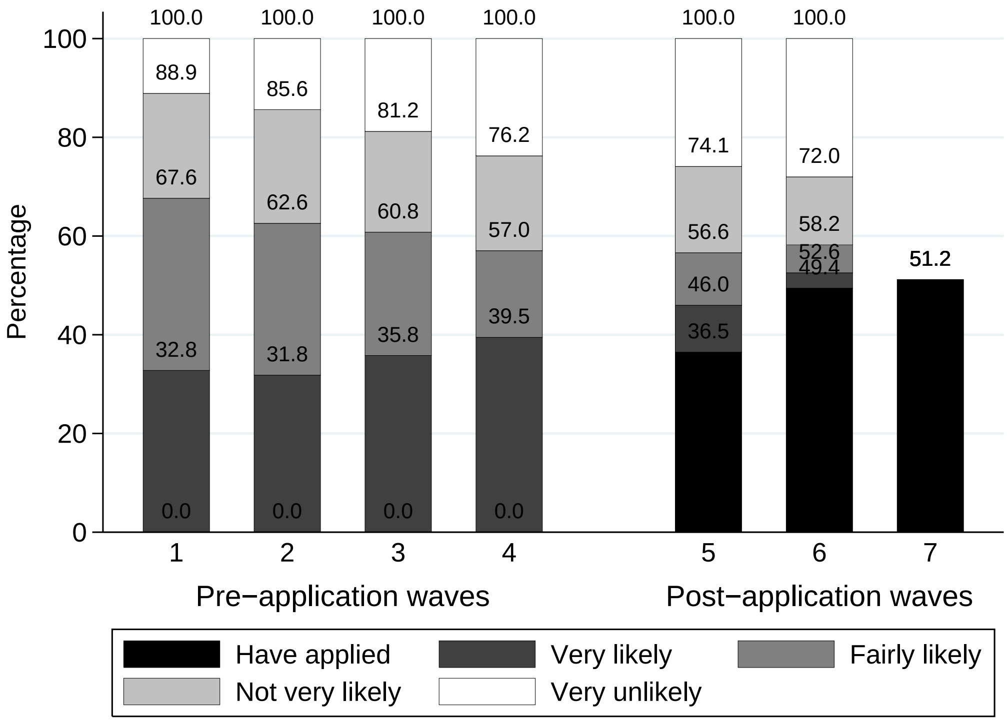
To get an initial impression of the evolution of young people's expectations during this period, Figure 1 shows for each wave, 1 to 7, the percentages of young people who report being 'very likely', 'fairly likely', 'not very likely' and 'not at all likely' to apply to university.[7] For the purposes of this graph, only individuals with expectations data throughout the survey are included (i.e. a balanced panel or complete case sample). However, as discussed above, this restriction is relaxed after this point. From Wave 5 onwards it is necessary to include an additional category for those who have actually applied. In Wave 7, only a measure of having actually applied to university by this point is reliably available. The overall percentage who are 'likely' (or who have already applied in later waves) can be seen by following the cumulative percentage above the 'fairly likely' blocks in Figure 1.

Overall, the proportion reporting that they are 'likely' to apply to university declines substantially from 68% in Wave 1 to 57% in Wave 4, at the end of the first year following GCSEs. There is essentially no change in Wave 5, when actual applications begin to be included (treated, for this purpose, as

[6]In colloquial English, the expression 'not very likely' means 'fairly unlikely', rather than its more literal interpretation of anything less than 'very likely'.

[7]Individuals may also respond that they 'don't know' whether they are likely to apply to university; however, this is not a common response (4.4% of weighted Wave 1 respondents) and I choose to classify those who report 'don't know' as being 'not very likely' to apply to university.

Figure 1: Young people's expectations of university application, Wave 1 (age 13-14) to Wave 7 (age 19-20)



Notes: Sample: Wave 7 respondents with non-missing data on university expectations and university application at each wave (complete case analysis). 'Don't know' (4.4% of weighted Wave 1 respondents) treated as 'not very likely'. Wave 7 attrition and non-response weights applied. Unweighted sample size = 8,029. Data labels show cumulative percentages.

'likely' to apply, given that they are effectively 'certain' to apply), before a small rise in Wave 6 when the study members would be completing any Further Education (two years of post-compulsory education). There is no reliable question on expectations of application to university in Wave 7, only a report of whether individuals have already applied. However, individuals will continue to enter university over the subsequent few years (or even later as mature students) (UCAS, 2012). It is therefore probable that a small percentage of the sample would have responded that they were likely to expect to apply to university if they had been asked in Wave 7.

In any case, as the aim of this paper is to understand changes in young people's expectations in the period leading up to making an application, the analysis in this paper is deliberately curtailed at the last wave in which individuals have not yet started applying to university (Wave 4, or roughly age 17). Analysing the period in which individuals apply to university would introduce bias from non-random movement of individuals out of the sample, caused by having actually made an application. I discuss this, along with other kinds of 'right-censoring' in Section 4.

For the analysis in this paper, I dichotomise the expectations variable into a distinction between young people who are 'likely' ('very likely' or 'fairly likely') or 'unlikely' ('not very likely' or 'not at all likely') to apply to university.[8] Assuming that young people are utility maximising (and that they give honest responses), they will report that they think it is likely that they will apply to university if they judge that the benefits they will derive from making an application exceed the costs they will experience as a result of doing so. They switch to thinking that it is unlikely that they will apply if their assessment of these costs and benefits changes to the point that the balance has shifted in the other direction. Many of the factors that will influence these decisions are not observed. However, I use those that are observed to assess which factors seem important in altering young people's perceptions of their potential to gain from higher education.

One problem with analysing expectations, rather than observed behaviour, is that 'talk is cheap'. This is an analysis of individual's stated preferences, rather than the revealed preferences indicated by their actions i.e. actually making an application to university. Cognitive biases, such as social desirability bias, may affect the responses. However, young people's reported expectations do seem informative as to the application behaviour observed in later waves of the LSYPE. 64% of those who say they think it is likely ('very' or 'fairly') that they will apply to university at age 14 have done so by the last point of observation (and more may do so at a later date), while only 22% of those who say

[8]Anders and Micklewright (2013) analyse the trends of those who report being 'very likely' to apply to university, finding that, unlike the overall proportion who report being 'likely', this in fact rises over time. This appears to be driven by a tendency for individuals' expectations to 'harden' over time, with those who report being 'fairly likely' tending towards reporting 'very likely', while those who report being 'not very likely' tend towards reporting 'not at all likely'.

they think it is unlikely have done so by the same time.

Use of a stated preference measure as an outcome variable in duration modelling in this way is innovative,[9] but raises some issues. The method is more normally employed to analyse transitions between clearly definable states, such as movement between employment and unemployment. Individuals' evaluation of their probability of applying to university will be subject to far more measurement error than transitions between such states. For example, factors such as an individual's bad mood on the day of the interview could tip them from reporting 'fairly likely' to reporting 'not very likely', if their general assessment of the costs and benefits of applying to university are finely balanced. Unlike in a standard binary regression model this does not just cause dependent variable measurement error. Since the sample for duration models depends on the reported expectation of application in the previous period, measurement error could also affect this. This will bias overall transition rates upwards, and may also affect estimated coefficients if groups are differentially affected by measurement error.

3.2 Sequences of expectations

To illustrate the form of data used in duration analysis, in Figure 2 I present the ten most common sequences of individuals' expectations between ages 14 and 17 observed in the dataset, which account for around 85% of the sample. Solid lines represent periods when the individual reports being likely to apply to university; dotted lines represent periods when individuals report being unlikely to apply to university; the absence of any line indicates missing data (including due to item non-response, unit non-response and attrition) at this time point. I have chosen to highlight the start and end of periods of being 'likely to apply': a vertical tail to the line represents the point at which the spell is observed to begin; and an arrowhead represents the point at which the spell is observed to end in a transition to the person reporting that they are 'unlikely to apply' to university.[10]

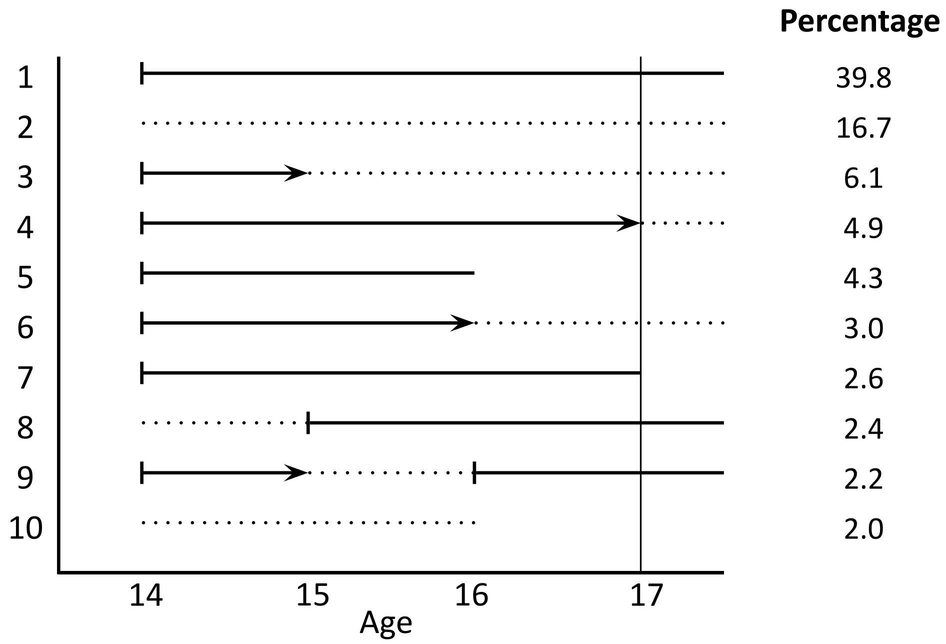
After exclusions, there are a theoretical maximum of 35 possible sequences of expectations during this period, all of which are observed in the data. The most frequent sequence of expectations (40% of the sample) is for individuals to report being 'likely to apply' at every interview from age 14 to age 17. The second most frequent (17% of the sample) is reporting being 'unlikely to apply' at every interview from age 14 to age 17.

To provide context to these records, in Table 1 I provide summary statistics about individuals who

[9]Some precedent is provided by studies of the dynamics of poverty (Bane and Ellwood, 1986, for example) where measurement of income may affect movement in or out of poverty.

[10]I could just as easily have highlighted the start and end points of periods of being 'unlikely to apply', but could not do both without loss of clarity.

Figure 2: Ten most common sequences of individuals' expectations from age 14 to 17 and the percentage of the total sample with each sequence



Notes: A solid line indicates that the individual reported they were 'very likely' or 'fairly likely' to apply to university at the most recent wave. A dotted line indicates that the individual reported that they were 'not very likely' or 'not at all likely' to apply to university at the most recent wave. The absence of a line indicates that there was no report from the individual at the most recent wave. An arrow tail at the start of a spell highlights that in the previous wave the negative outcome was observed. An arrow head at the end of a spell highlights that in the following wave a negative outcome was observed. The vertical line at age 17 highlights that this is the final point of observation and hence data beyond this point only provide information on whether the spell was censored (whether by no change or missing data) at this point. Calculation of frequency of spell types was weighted using LSYPE Wave 2 attrition and non-response weights. Individuals with missing data in either of Waves 1 or 2 are excluded. Percentages based on total sample size of 11,249.

Table 1: Summary statistics about sequences of expectations

Group	N	Percentage	SES Index
1	4,503	40.2	0.45
2	1,857	16.6	-0.49
3	673	6.0	-0.35
4	547	4.9	-0.07
5	478	4.3	-0.23
6	342	3.1	0.04
7	279	2.5	-0.04
8	269	2.4	-0.53
9	249	2.2	0.05
10	225	2.0	-0.27
Other	1,828	15.9	-0.30
All	11,249	100	0.00

Notes: Adjusted using LSYPE-provided Wave 2 survey design, attrition and non-response weights. Individuals with missing data in either of Waves 1 or 2 are excluded.

have the sequences of spells in Figure 2. I also include a category for all remaining groups, which makes up about 16% of the sample and is somewhat less advantaged than the average individual. The SES index (discussed further in Section 3.3) is standardised such that the sample mean is 0 and the standard deviation is 1. Individuals who always report being likely to apply to university (type 1) are, on average, half a standard deviation more advantaged than the sample as a whole. Conversely, those who always report being unlikely to apply (type 2) are roughly the same amount less advantaged than the sample as a whole.

Another important feature of the data is that, although an individual's changes in expectations seem more likely to be a continuous underlying process, I only observe their reported expectations in surveys once a year. This is, therefore, 'discrete time', as opposed to 'continuous time', data. This is illustrated in Figure 2: spells only start or end at exact ages, never somewhere in between. It follows that the models in this paper estimate the probability of transition between these observation times, rather than at any arbitrary time point. A further limitation of discrete time data is that some transitions back and forth between the observation points are hidden, which may bias overall transition rates downwards. The issues arising from use of discrete time data in duration modelling are discussed further in Section 4.

3.3 Measurement of SES

The LSYPE includes a rich set of data on participants' characteristics. These will be important in measuring young people's socio-economic status (SES) well, in order to assess its association with changes in their expectations of applying to university. Household income, parental education, and parental occupational status are all important in measuring SES (Hauser, 1994). The rich data will also be important in controlling for other factors correlated with SES, but which seem likely to make an important contribution in their own right, such as demographic characteristics, school characteristics, local area, and prior academic attainment. I return to these in the following section (Section 3.4).

Household income is measured at each wave between 1 and 4. As the method used to collect information on income varies somewhat from wave to wave and previous research has suggested 'permanent' income (rather than transitory income) has a much larger effect on young people's educational outcomes (Jenkins and Schluter, 2002, p.2), I construct an approximation of the household's 'permanent' income by averaging across the four measures. I also equivalise my income measure by dividing it by the square root of household size, thus recognising the importance of family resources being stretched further in larger households. Household income is underestimated to some extent in the

LSYPE, relative to other social surveys where it is a major focus (Anders, 2012a).

Parental education seems likely to play a role in the formation of young people's educational expectations (Ganzach, 2000), not least because young people whose parents went to university are more likely to see it as a natural next step in their education. Indeed, Table 3 shows that, at least based on the initial report of expectations at age 14, more of the young people who report that they are 'likely to apply' to university have at least one parent who themselves received higher education than young people who report that they are 'unlikely to apply'. Data on parental education is collected from both parents (where available) at each wave between 1 and 4 using the same questions; where both parents' education level are recorded and these differ I use the highest. Unsurprisingly, there is very little change over time, since most parents have already completed the highest educational level they will achieve by this stage of their lives.

Parents' occupational status is recorded in the LSYPE using the National Statistics Socio-Economic Classification (NS-SEC), which was designed to capture social class differences between the different occupational types (Rose and Pevalin, 2001). It is based on questions about job title, role and responsibilities asked of both parents (where available) at each wave between 1 and 4. As with parental education, where both parents' occupational status are recorded I use the highest, and, also as with parental education, there is little change in this variable over the period of analysis. I collapse the classification into four ordinal groups[11]: managerial and professional occupations; intermediate occupations; routine and manual occupations; and long-term unemployed.[12] Social class is seen by sociologists as a key element of an individual's SES, as "the experience of individuals in terms of economic security, stability and prospects will typically differ with the class positions that they hold" (Goldthorpe and McKnight, 2004). Particularly relating to the purposes of this paper, sociological theory suggests that "young people (and their families) have, as their major educational goal, the acquisition of a level of education that will allow them to attain a class position at least as good as that of their family of origin" (Breen and Yaish, 2006, p.232). This implies that individuals from different class backgrounds will have, on average, different educational expectations.

I combine the above measures of household equivalised 'permanent' income, highest parental education, and highest parental occupational status into a single index of SES.[13] This provides a broader measure of family circumstances that any one measure would provide. I use principal components

[11]Some sociologists are critical of attempts to express social class in ordinal terms, most particularly in how self-employed individuals should fit into such a hierarchy (Rose et al., 2005).

[12]Individuals experiencing short-term unemployment at the time of interview are allocated a group based on their most recent job.

[13]All measures from age 14 (except income, which is averaged over available observations between age 14-17), except where not available due to item non-response at age 14, when data from later in the survey was used.

analysis with a polychoric correlation matrix (Olsson, 1979; Kolenikov and Angeles, 2009) to construct a single index, which explains roughly three quarters of the variation in the three individual measures.[14] I divide individuals into quintile groups on the basis of this SES index; Table 2 reports the family characteristics of the median individual in each quintile group, demonstrating increasing SES across all three dimensions, as would be expected.

Table 2: Median family characteristics by quintile group of socioeconomic status index

Quintile group	Q1	Q2	Q3	Q4	Q5
Parental Education	< A*-C GCSE	A*-C GCSE	A Level	HE < Degree	Degree
Occupational Status	Routine occupations	Routine occupations	Intermediate occupations	Higher occupations	Higher occupations
Family Income (£p.a.)	5,699	9,549	12,992	16,433	29,941
N	2,585	2,221	2,171	2,201	2,071

Notes: Adjusted using LSYPE-provided Wave 2 survey design, attrition and non-response weights. Standard errors, clustered by school, in parentheses. Family income is equivalised by dividing by the square root of household size. Sample: Wave 2 respondents with non-missing data on university expectations ('don't know' treated as 'not very likely') and university applications.

3.4 Measurement of other factors

The dataset also includes a rich set of participant characteristics and experiences. As discussed in Section 2, many of these factors are correlated with SES. However, they may also have independent effects of their own, with their exclusion resulting in omitted variable bias. It follows that it is important to be able to control well for these other factors to isolate the influence of SES. In this section I discuss the measurement and importance of academic ability, demographic characteristics (age, gender and ethnicity), school characteristics, traumatic events, and local labour market conditions.

One of the advantages of duration modelling is that it allows me to take into account different values of explanatory variables at different times. As such, in addition to describing potential explanatory factors in the dataset, I also assess their potential use as valid time-varying covariates. This requires that they are measured repeatedly and consistently throughout the LSYPE, since measurement in differing ways might result in changes that are not due to any underlying change in circumstances. Box-Steffensmeier and Jones (2004, p.110-112) also highlight the importance of understanding the temporal ordering of time-varying covariates and the events it is being claimed that they are caus-

[14]Despite the presence of non-continuous variables, constructing my SES index using any of the following alternative methods makes no substantive difference (correlation coefficients between the indices $r > 0.98$) to my SES quintile groups: principal components analysis applied to a Pearson's correlation matrix; factor analysis treating the income, education and occupational status as continuous and using full information maximum likelihood (FIML) to deal with missing data; factor analysis treating income as continuous, and education and occupational status as ordinal, using FIML, but no weights. Given this, I am confident that my SES index is robust.

ing. Since, by their nature, time-varying covariates are not fixed, it is particularly important to assess whether, in this case, such covariates are plausibly being affected by changes in young people's expectations of applying to university. This eventuality, referred to as reverse causation, would result in endogeneity bias to the estimates (Goodliffe, 2003).

Table 3: Summary statistics of sample by whether young person reports being likely or unlikely to apply to university at age 14

Variable	Mean of Unlikely	Mean of Likely	Mean of Whole Sample	Standard Deviation
SES Index (Z-Score)	-0.40	0.20	0.00	1.00
	(0.02)	(0.02)	(0.02)	
Equivalised Family Permanent Income	12464.07	18029.33	16199.21	12220.12
	(209.35)	(256.24)	(208.44)	
At least one parent has Higher Education	0.06	0.25	0.19	0.39
	(0.00)	(0.01)	(0.01)	
At least one parent has 'Higher' Occ. Status	0.26	0.49	0.41	0.49
	(0.01)	(0.01)	(0.01)	
Lone Parent	0.28	0.20	0.22	0.42
	(0.01)	(0.01)	(0.00)	
Gender: Male	0.55	0.48	0.51	0.50
	(0.01)	(0.01)	(0.01)	
Ethnicity: Non-White	0.07	0.16	0.13	0.34
	(0.00)	(0.01)	(0.01)	
Age 11 Attainment Z-Score	-0.48	0.23	-0.00	0.97
	(0.02)	(0.02)	(0.02)	
Age 16 Attainment Z-Score	-0.60	0.29	-0.00	1.00
	(0.03)	(0.02)	(0.02)	
Attend Independent School	0.02	0.10	0.07	0.26
	(0.01)	(0.01)	(0.01)	
Attend Grammar School	0.01	0.05	0.04	0.19
	(0.00)	(0.01)	(0.01)	
Attend school with Sixth Form	0.52	0.56	0.55	0.50
	(0.02)	(0.02)	(0.02)	
Local Unemployment Rate (%) at Age 14	4.61	4.80	4.74	2.14
	(0.07)	(0.07)	(0.06)	
N	3686	7523	11209	

Notes: Weighted using LSYPE Wave 2 sample design and non-response weighted weights. Standard errors, clustered by school, in parentheses. Household income is equivalised by dividing by the square room of household size.

Correlation between academic ability and SES would lead to upward biased estimates of the effect of SES on young people's expectations of attending university, if it is not included in the model. Academic attainment provides an imperfect proxy for the unmeasurable individual trait of ability. A particularly important imperfection is that SES is likely to have an effect on the attainment measures available in the LSYPE. This suggests that models including attainment may underestimate the influence of SES. The LSYPE provides measures of academic attainment through linkage to selected elements of the National Pupil Database (NPD). This provides information on the young people's academic attainment from Key Stage 2 (age 11), Key Stage 3 (age 14) and Key Stage 4 (age 16). Having high-quality, seldom-missing data on prior attainment is a major advantage compared to many surveys. Key Stage 5 data (from qualifications taken at ages 17 and 18) are now available as part of the

LSYPE release. However, I do not use them as part of this analysis, since the relevant examinations are taken after the period of this analysis.

Some of the academic attainment data from ages 11 and 14 are missing where an individual was not in the state education sector and hence either did not take the relevant tests (SATS) or, if they did, the school chose not to report them. Pupils at independent schools are under no obligation to do either, although many do. A missing variable dummy is employed for Key Stage 2 scores to prevent these individuals from being excluded from my analyses. This is not an option for Key Stage 3, since the missing variable dummy would be almost perfectly collinear with an indicator of independent school attendance. Given this problem, the fact that children are unlikely to change schools immediately after taking their Key Stage 3 SATS and the low stakes nature of Key Stage 3 SATS I decide not to include it in my analysis.[15]

For Key Stage 2 (KS2), I use the average raw point score across all three subjects (Maths, English and Science[16]). KS2 SATS are relatively low stakes examinations for pupils, although they are rather higher stakes for primary schools and there is some limited use by secondary schools for tasks such as sorting pupils into ability groups. After weighting, there is a roughly normal distribution of scores ranging between approximately 0 and 100. The mean score is 65.5 and the median individual obtains a score of 67.3. I standardise this variable, creating a 'Z-score' with a mean score of zero and a standard deviation of one.

For Key Stage 4 (KS4), I use the official capped GCSE score. GCSEs (General Certificates of Secondary Education) are high stakes public examinations, taken at the end of compulsory education. They potentially have a large bearing on the individual's future education and/or employment. After weighting, the capped point score gives a range of scores from 0 to 483, with a mean of 306 and a median of 326. The capped point score is calculated from an individual's best 8 GCSEs or equivalent qualifications. This is in contrast to the uncapped score, which uses all GCSEs and equivalents taken and hence is more subject to manipulation by schools. Again, I standardise this so that the score has mean zero and standard deviation one. However, it should be noted that there is some potential for reverse causation in the relationship between KS4 performance and young people's educational expectations, in that individuals' beliefs about their likelihood of applying to university may affect the effort they put into these examinations.

The LSYPE collects data on young people's demographic characteristics, including their gender, age

[15]It is also worth noting that Key Stage 3 SATS were abolished in England in 2008 (BBC News).

[16]In the raw scores, Science is out of 80. I rescale it to be out of 100, ensuring it receives the same weight as Maths and English.

and ethnicity. While neither gender nor age are likely to be correlated with SES, they are both likely to be important in explaining changes in young people's expectations.[17] However, individuals with different ethnicities have, on average, different levels of SES (Strand, 2014). As such, failure to control for ethnicity may result in effects stemming from, for example, cultural differences between ethnicities, being incorrectly identified as SES effects. In the LSYPE, ethnicity is initially collected according to young people's self-designation, and classified into the groups White, Mixed, Indian, Pakistani, Bangladeshi, Black Caribbean, Black African and Other before the data are released.

The input of schools and teachers is important in shaping young people's educational choices. For example, Alcott (2013b) finds evidence that teacher encouragement makes it more likely that young people remain in education past the minimum leaving age. Likewise, Sanders et al. (2013) report that within-school provision of information on university increases stated likelihood of application. The LSYPE includes data on the young person's school type at time of sampling. Of particular interest, this allows me to identify those who attend academically selective 'grammar' schools (4% of the age 14 sample) and those who attend fee-paying independent schools (5% of the age 14 sample). Table 3 shows that a significantly larger proportion of those who think it likely that they will apply to university at age 14 than those who think it is unlikely are in one of these types of schools. It is also the case that individuals from higher SES backgrounds are more likely to be in such schools. It is not clear how much of the influence of schools is an 'independent' effect and how much reflects SES bias in the intake of different types of school. As such, in the same way as was discussed above regarding inclusion of prior attainment in a model, conditioning on school characteristics may result in an underestimate of the total influence of SES.

Traumatic events within a family, such as job loss, separation or bereavement, might also be expected to have a negative influence on young people's educational expectations. Such events are to some extent random and, hence, effects would be at least partly independent of those of SES. However, there is likely to be some correlation.

The employment status of parents in the household are recorded at each wave. Drawing on previous evidence that finds an association between even short periods of worklessness and lower educational expectations (although these do not persist when additional controls are added) (Schoon et al., 2012, p.38-39), I construct a cumulative indicator of whether the young person has experienced being in a workless household by the time of each wave's interview. As I do not have data before age 14, it is not possible for this to include periods of worklessness before this point. Nevertheless, 22% of the young

[17]Given the relationship between age and the passage of time in this dataset, I discuss the inclusion of age in the models further in Section 4.

people's parents (after weighting) reported neither parent being in work in at least one wave. I judge that it is unlikely that young people's educational expectations affect changes in employment status in their household, and hence the risk of endogeneity bias is low. However, sociologists emphasis that an important element of social class is the increased economic security of those with higher SES (Goldthorpe and McKnight, 2004, p.6). Once again this implies that, once this factor is controlled for, my estimates of the influence of socio-economic status are likely to be understated.

I use information on the marital status of the 'main parent'[18] in a similar way as the employment indicators, constructing a cumulative indicator of whether the young person has experienced this parent going through some kind of separation (including bereavement) up to the point of each wave's interview. Unlike with the indicator for workless households, retrospective questions (asked at the first wave of the survey) about relevant events since the young person was born mean that this does cover the period before age 14. 28% of young people's main parents report having experienced such an event by the final interview with them. I define a cumulative measure on the grounds that negative consequences on a young person's attitudes from such an event are unlikely to be limited to one year. Again, I judge that there is unlikely to be problems of reverse causation with this time-varying covariate.

Local labour market conditions are important in predicting young people's decision to apply to university: other things being equal, individuals who face circumstances in which the labour market looks less promising are more likely to remain in education longer (Reynolds and Pemberton, 2001; Fumagalli, 2012). However, on average, SES and worse local labour market conditions are likely to be negatively correlated. Unlike with the characteristics discussed above, this implies that not including this factor in the model may understate the impact of SES. To include this in my models I make use of data on the Local Authority (LA) area in which the young person's home is located is also available from the LSYPE. I use this LA identifier to link this with data on unemployment in the local labour market[19] from the Annual Population Survey (Office for National Statistics, 2004, for example). I use the unemployment rate for those aged 16-64 in the individual's LA area, with separate figures for males and females. In a small number of LAs the figures are suppressed, due to small numbers in the data. In such cases I use the Government Office Region unemployment rate (or in extremis the national unemployment rate) to avoid missing data.

[18]Defined as the parent most involved in the young person's education. Where there is only one parent in the household they are, by definition, the main parent.

[19]Since the aim is to capture the labour market conditions individuals face, it would be better to use areas designed to reflect this. Local Authorities do not necessarily reflect local labour markets well, especially in larger, rural authorities. A better alternative would be Travel To Work Areas (TTWAs). Unfortunately, information that would allow me to identify in which TTWA an individual resides is not available in the LSYPE general release.

4 Duration modelling

Duration modelling, also known as survival analysis or event history analysis, is not a common technique in educational research (Alcott, 2013a, p.50-51). However, it has several key features that make it a useful tool to address the question of changes in young people's expectations, specifically models of change i) from 'likely to apply' to 'unlikely to apply' and ii) from 'unlikely to apply' to 'likely to apply'. In this section, I introduce its key features, concepts and their importance for the application in this paper.

Central to duration modelling is the concept of the 'spell'. A spell is an uninterrupted period of time during which a given individual remains in the same state; in this case, consistently reporting that they are 'likely to apply' to university, or conversely, consistently reporting that they are 'unlikely to apply'. Figure 2 shows spells as uninterrupted periods as solid lines ('likely to apply') or dotted lines ('unlikely to apply'). In some applications of duration modelling the end of a spell is permanent (or effectively permanent), such as in models of an individual's death after the onset of a disease. However, in this application individuals can report being 'likely to apply', then 'unlikely to apply', and then 'likely to apply' again.[20]

Since participants can move back and forth between being 'likely' and 'unlikely', the same individuals may appear in both sets of models at different time points. One can see that this is indeed the case by calculating the proportion of the sample that ever report being 'likely to apply' to university and the proportion that ever report being 'unlikely to apply'. First, considering the transition from 'likely to unlikely', 79% of the Wave 2 weighted sample (representing 9,247 out of 11,249 individuals before weighting) in the dataset report being 'likely to apply' to university (and, hence, are ever in a position to make a transition to being 'unlikely to apply') in at least one wave. In the other direction, 52% of the Wave 2 weighted sample (representing 5,330 out of 11,249 individuals before weighting) report they are 'unlikely to apply' (and, hence, are ever in a position to make a transition to being 'likely to apply') in at least one wave. In total, this sums to 131% of the sample, demonstrating the significant overlap. One can also see this is the case by looking at the sequences of expectations observed in the data in Figure 2: individuals of type 3 are included in the model of 'likely to unlikely' at age 15, then in the model of 'unlikely to likely' at ages 16 and 17.

To highlight the implications of using duration modelling, relative to a model of differences between the start and the end of the time period under consideration, in Table 4 I compare the proportion of

[20]It should be noted that one reason for such sequences of transitions could be measurement error. This makes allowing for multiple spells particularly important, since ignoring spells after the first would compound the error.

Table 4: Proportion of young people saying they are likely or very likely to apply to university - always reported likely vs. current wave

Wave	Always likely	Current wave
1	0.676	0.676
2	0.552	0.626
3	0.484	0.608
4	0.429	0.570
5	0.399	0.566
6	0.384	0.582

Notes: Analysis weighted using LSYPE Wave 7 design and non-response weights. Sample: Wave 7 respondents with non-missing data on university expectations ('don't know' treated as 'not very likely') and university applications. Unweighted sample size = 8029. 'Always likely' column reports proportion of the sample who have always reported being 'very likely' or 'fairly likely' to apply to university up to and including the wave in question. 'Current wave' column reports the simple proportion of the sample who report being 'very likely' or 'fairly likely' to apply at the wave in question.

individuals who at all points up to and including the relevant wave have reported that they think it 'likely' that they will apply to university (in the left hand column), with the proportion who think it is 'likely' that they will apply at that particular point in time (in the right hand column). As also noted in Figure 1 earlier, the proportion who think it is 'likely' that they will apply at a given point in time falls from 68% at Wave 1 to 57% by Wave 4. However, the reduction in those who have always reported being likely to apply is much greater: from 68% at Wave 1 to 42% by Wave 4. This difference is caused by individuals who start reporting being 'likely to apply' after Wave 1 (e.g. individuals of type 8) in Figure 2.

The larger reductions in the proportion who have always reported being 'likely to apply' demonstrates the additional information on transitions that is picked up by using this approach. This information would be ignored if I only modelled the difference between the start and the end of the time period under consideration. In fact, as I allow for multiple transitions, the differences are even larger than suggested in this table, since the analysis in this paper recognises that individuals can, in principle, switch back and forth as many times as there are observation periods (e.g. individuals of type 9 in Figure 2). Each transition from being 'likely to apply' to being 'unlikely to apply', even multiple transitions by the same individual, is captured as part of the modelling.

My multiple regression-based duration models will allow for multiple spells in a state, since this is preferable to concentrating only on the first one. However, my modelling strategy treats multiple spells as being independent from one another, making the assumption that there is no causal effect of one spell on any later spells (either of the same type i.e. 'likely to unlikely', or the converse transition i.e. 'unlikely to likely').[21]

[21]However, see discussion of clustering of standard errors in Section 6. Furthermore, I attempt to partially relax the assumption of independence of multiple spells of the same type using random effects models, discussed in Appendix C. However, it maintains the assumption of no effect of an individual's spell of being 'likely to apply' on subsequent spells

The passage of time is, as the name suggests, fundamental to duration modelling. Models can include the length of time an individual has spent in a spell before making a transition, not throwing away this considerable amount of information as would be done in a traditional binary choice model (DesJardins, 2003; Box-Steffensmeier and Jones, 2004; Jenkins, 2004). However, as individuals in the LSYPE are all (approximately) the same age at the same point in time, where spells begin at the same point it is impossible to distinguish between age and duration effects. In the data, some spells do start at different time points, but there is not enough variation to disentangle the effects of age and duration. At this stage of life age effects are more important to educational expectations than duration in the state, and concentrate on these. Other important characteristics of individuals may also change over time and duration modelling is able to incorporate such time-varying covariates[22]

As discussed in Section 3, since I have discrete (as opposed to continuous) time data, I use discrete time duration modelling techniques, as the most appropriate. One potential problem with this is that, since young people are born in different months and the LSYPE interviews are staggered over several months, there will be some variation in individuals' age by month when they are give their responses. In order to reduce the possibility that this could affect results, I include individuals' month of birth and month of interview in all my regression models, attempting to standardise results as if individuals were all both born and interviewed in August each year.

A key concept in duration modelling is that of an individual being 'at risk' of making a transition, and therefore relevant to my modelling. When modelling a transition it only makes sense to consider those who are in a position to make that transition. As a minimum, this excludes those who already in the state of interest. For example, it does not make sense to consider the probability that someone who already reports being 'unlikely to apply' to university *becomes* 'unlikely to apply' to university. While it may be interesting to consider the question of whether an individual *remains* 'unlikely to apply', that is a different question (and, in fact, just the inverse of my other model: whether an individual currently reporting being 'unlikely to apply' becomes 'likely to apply'). In some applications individuals may become not at risk in other ways.

Duration modelling can also treat expectations data that are missing as 'censored', rather than dropping individuals for whom expectations are not observed (even in only one wave) from the sample. 'Censoring'[23] is where the start and/or end points of a spell is not observed in the data. It has the consequence that the true length of the spell is unknown, only that it is at least as long as the period

of being 'unlikely to apply'.

[22]This was discussed further in Section 3.4.

[23]Censoring is sometimes confused with 'truncation'. This is when the probability of inclusion of a spell is affected by its length or where spells are cut short for the same reason. I do not have to deal with truncation in my data.

it is observed to last.

When the start of a spell is not observed this is referred to as 'left censoring'; this can be particularly problematic, as it prevents modelling of duration dependence, since one does not know how long a spell has lasted at any given point (Iceland, 1997). However, as discussed in Section 3.1, I treat all spells as starting at age 14 and, hence, exclude the possibility of left censoring in this dataset by construction.

Not observing the end of a spell is referred to as 'right censoring'. Taking the example of models for the 'likely to unlikely' transition, this occurs where 'likely to apply' is observed in the final report for an individual, whether this is due to the end of the period under analysis (at age 17 in this case), or earlier as a result of attrition. Still concentrating on the 'likely to unlikely' transition, there is right censoring in the sequences of spells in Figure 2 for individuals of type 1, 8, and 9 (in the case of the final observation being still 'likely to apply'); and types 5 and 7 (resulting from attrition).

Treating individuals who attrit from the sample as right censored will only result in unbiased estimates under the assumption that this missing data censoring is 'uninformative' (Clark et al., 2003, p.236), i.e. that individuals whose outcomes are missing are just as likely to make a transition between reporting being 'likely to apply' to university and being 'unlikely to apply' (or vice versa) as the individuals that are observed. It seems unlikely that this assumption is justified. However, van den Berg et al. (2006) suggests it is likely that while informative attrition will affect the rate of transitions, it is less likely to bias the effect of covariates on those rates. As a robustness check, I also repeat my analysis including only those still participating in the survey at Wave 4 (when the response rate relative to Wave 1 has fallen to 73% (Collingwood et al., 2010, p.52)), using the LSYPE-provided attrition and non-response weights for Wave 4.[24]

All of these features are important in fitting the most appropriate model to understand changes to young people's expectations during these critical years for their education.

5 Nonparametric analysis of transitions

In this paper I model the probability and timing of young people's transitions from reporting they are 1) 'likely to apply' to 'unlikely to apply' or, conversely, 2) 'unlikely to apply' to 'likely to apply'. Restricting my attention to those who are 'at risk' of making each transition, it follows that I am interested in the likelihood of the following events:

1. for the transition from 'likely to apply' to 'unlikely to apply': whether individuals, who at the

[24] I report the results of this analysis and discuss the differences in Appendix B.

previous wave said they were 'likely to apply' to university, switch to reporting that they are 'unlikely to apply'; and

2. for the transition from 'unlikely to likely': whether individuals, who at the previous wave said they were 'unlikely to apply' to university, switch to reporting that they are 'likely to apply'.

To begin exploring these transitions, I conduct non-parametric analysis of the probability and timings of transitions between being 'likely' and 'unlikely' to apply to university and consider the association between the probability of making a transition and young people's SES. In order to do this I make use of Kaplan-Meier estimates of the probability that spells have not ended with a transition by a given age. To obtain Kaplan-Meier estimates one first calculates, at each time point in the data, the number of individuals that do not make a transition divided by the number that are in a position to make a transition. The estimate for each time point is the product of all of the proportions just calculated from the first time point up to the time point in question. Kaplan-Meier estimates are able to handle right-censoring in the data, since individuals who are censored are removed from the denominator, since they are no longer 'at risk'. These estimates of 'survival' will be calculated both for the sample as a whole, and for sub-samples defined by SES.

In order to perform this analysis, I restrict the spells under consideration to those beginning at age 14 (the start of the dataset). By definition, this also means concentrating on an individual's first spell at risk, ignoring any later spells either as 'likely' or 'unlikely'. Below, I indicate the kinds of spells excluded as a result. Among the costs and benefits of the multiple regression-based analysis introduced in Section 6, this restriction will be relaxed.

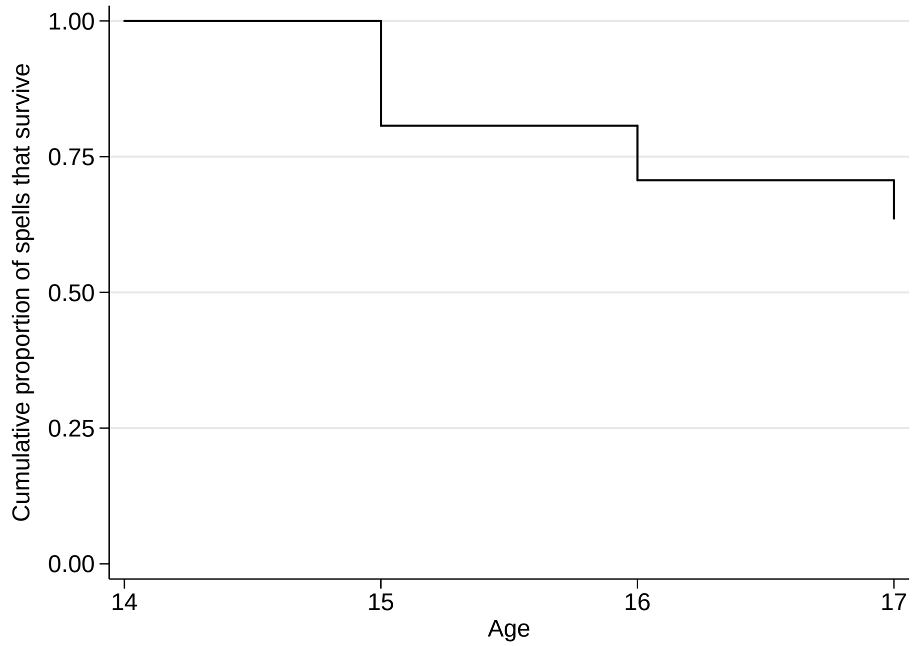
It was not possible to perform non-parametric statistical inference on the difference between estimated survival functions as part of this analysis. The relevant statistical test, the log-rank test, is "not appropriate" with sampling weights (StataCorp, 2013, p.446). Instead, I perform Cox regression-based tests, which make the proportional hazards assumption. However, I checked the robustness of this approach by performing log-rank tests of the equality of the survival curves estimated using unweighted data. In all cases the two sets of results were in agreement.

I first consider the transition from 'likely to unlikely', before moving on to the transition from 'unlikely to likely'.

5.1 From likely to unlikely

I begin by analysing the age at which young people stop thinking they are likely to apply to university. Relating this to the sequences of expectations shown in Figure 2, this means including the first (or only) spell of individuals of type 1, 3, 4, 5, 6, 7 or 9 (amongst others not shown in the diagram), but not the spell that type 8 spends reporting being 'likely to apply'. Nevertheless, this includes over 70% of the individuals in the data, with much of the remainder being individuals who never report being 'likely to apply' rather than individuals who are excluded simply because of this restriction.

Figure 3: Probability that an individual who reports being 'likely to apply' at age 14 has not moved to reporting that they are 'unlikely to apply', by age

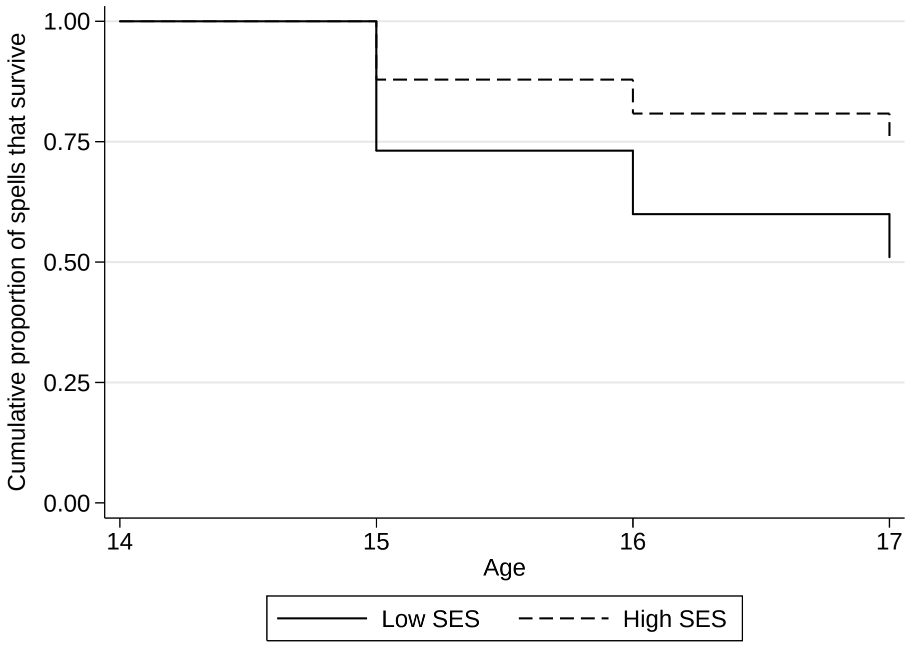


Notes: Kaplan-Meier estimated survivor function. Excludes spells beginning after age 14. Analysis weighted using Wave 2 sample design and non-response weights. Unweighted number of subjects: 6,129; weighted number of subjects: 6,009.

Figure 3 shows that 70% of periods of reporting being 'likely to apply' continue until at least age 16, at which point young people will be in the process of taking their GCSEs. Conversely, this means that 30% of such periods have ended with the individual switching to reporting they are 'unlikely to apply' by this age. Looking right to the end of the ages under consideration, roughly a third of the observed periods of being 'likely to apply' end by age 17. There are evidently a significant number of transitions during this stage of life. However, this sheds no light on the reasons for these changes, other than young people's age increasing.

A simple way of assessing the association between the probability of transition and family background is by estimating the survivor function for different groups of SES. For ease of interpretation I dichotomise SES into 'high' (comprising the top 40% of the distribution of my SES index) and 'low' (comprising the bottom 60% of the distribution). Figure 4 shows that individuals from lower SES households are more likely to make a transition to reporting 'unlikely to apply' than their richer counterparts throughout the period under analysis: 40% of those from lower SES backgrounds have made

Figure 4: Probability that an individual who reports being 'likely to apply' at age 14 has not moved to reporting that they are 'unlikely to apply', by age and household SES



Notes: Kaplan-Meier estimated survivor function. Excludes spells beginning after age 14. Analysis weighted using Wave 2 sample design and non-response weights. 'High SES' denotes individuals in the top two quintiles of SES, while 'low SES' refers to all other individuals. Unweighted number of subjects: 6,129; weighted number of subjects: 6,009. Cox regression-based test for equality of survivor functions rejects the null hypothesis of no difference (p<0.01)

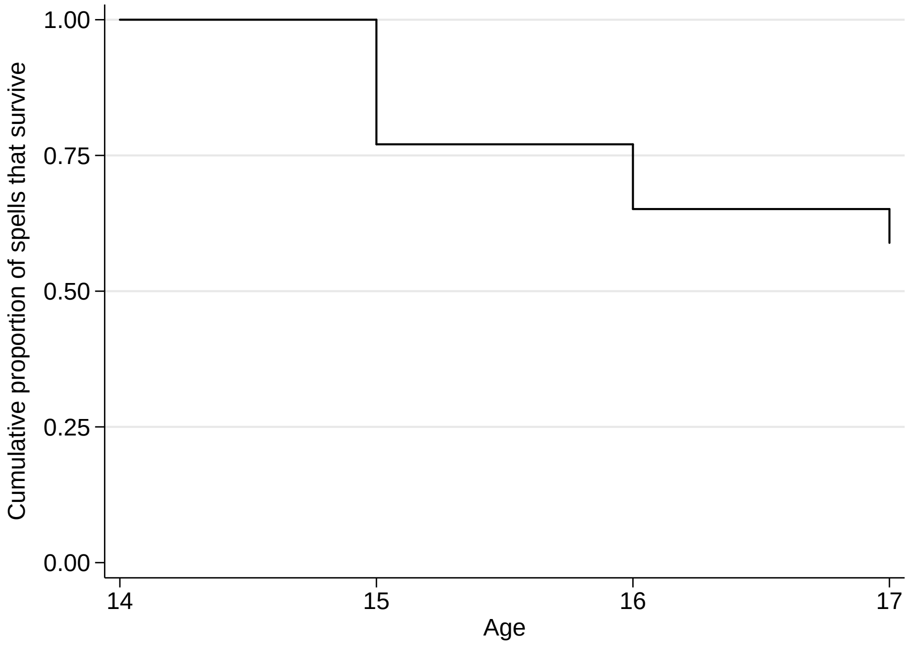
a transition from 'likely to unlikely' by age 16, whereas only 20% of those from high SES backgrounds have done so. Making the assumption of proportional hazards allows me to carry out a Cox-regression based test, which rejects the null hypothesis of no difference between the two estimated survivor functions (p=0.00).

5.2 From unlikely to likely

It is possible that the relationship between SES and young people raising their expectations is quite different from that associated with movement in the opposite direction. The analysis of this transition from 'unlikely to likely' includes the first (or only) spell from individuals of types 2, 8 and 10 in Figure 2, but not the spell that types 3, 4, 6 and 9 spend reporting being 'unlikely to apply'. This represents over 20% of the overall sample, but much of the remainder again comprises individuals who never report being 'unlikely to apply', rather than exclusions because of restricting to spells that start at age 14.

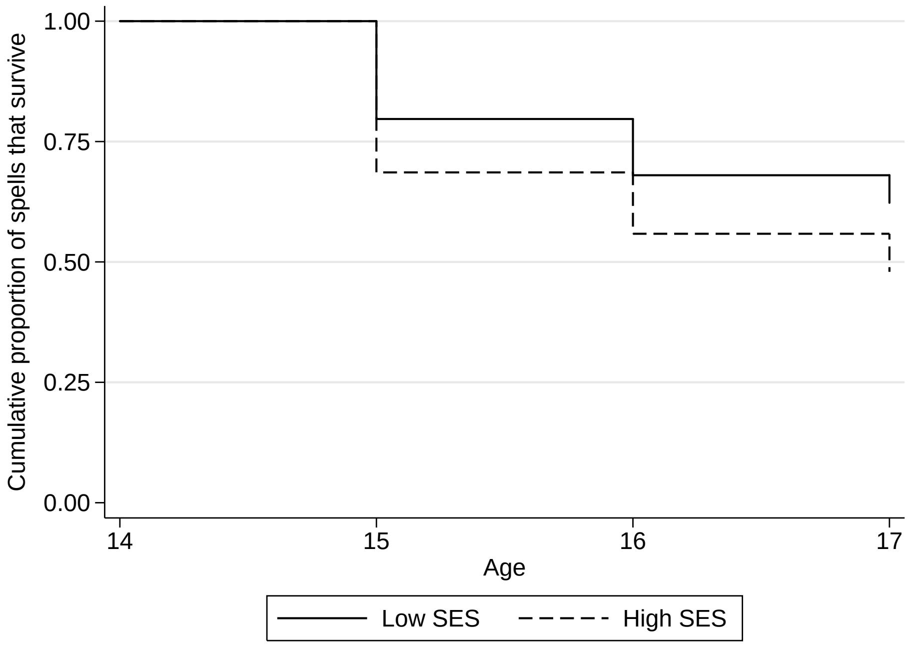
As with the opposite transition, Figure 5 shows the proportion of periods of being 'unlikely to apply' that do not end in transition to being 'likely to apply' by a given age. Almost 25% of spells end by age 15 and around a third of spells have ended in transition by the last point of observation at age 17.

Figure 5: Probability that an individual who reports being 'unlikely to apply' at age 14 has not moved to reporting that they are 'likely to apply', by age



Notes: Kaplan-Meier estimated survivor function. Excludes spells beginning after age 14. Analysis weighted using Wave 2 sample design and non-response weights. Unweighted number of subjects: 2,556; weighted number of subjects: 2,946.

Figure 6: Probability that an individual who reports being 'unlikely to apply' at age 14 has not moved to reporting that they are 'likely to apply', by age and SES



Notes: Kaplan-Meier estimated survivor function. Excludes spells beginning after age 14. Analysis weighted using Wave 2 sample design and non-response weights. 'High SES' denotes individuals in the top two quintiles of SES, while 'low SES' refers to all other individuals. Unweighted number of subjects: 2,556; weighted number of subjects: 2,946. Cox regression-based test for equality of survivor functions rejects the null hypothesis of no difference (p<0.01).

These are higher rates of transition than those seen for the same time points in my analysis of the transition from 'likely to unlikely' above, this despite a larger overall shift in the opposite direction. Although this initially seems counterintuitive, it is consistent because of the larger absolute numbers of young people who start out saying they are 'likely to apply' (as shown in Figure 1). Furthermore, it again highlights the large number of transitions between the two states.

In common with transitions from 'likely to unlikely', Figure 6 shows that there are clear socio-economic differences in the expected proportion of transitions from being 'unlikely to apply' to being 'likely to apply'. However, in this case those from the less advantaged groups are less likely to make a transition out of being 'unlikely' than their more advantaged peers. Again, a Cox regression-based test allows me to reject the null hypothesis of no difference between the two survivor functions (p=0.00).

Comparing Figure 6 with Figure 4 it is clear that the differences in rates of transition from being 'unlikely' to being 'likely' by SES are markedly smaller than for the transition in the opposite direction: by age 16 68% of those from lower SES backgrounds have made a transition from 'unlikely to likely', while 56% of those from more advantaged backgrounds had done so. This suggests that more of the inequality in expectations builds from less advantaged individuals having a higher probability of switching to reporting being 'unlikely', than from movements in the other direction. Nevertheless, the inequality in probability of transition from 'unlikely to likely' compounds the widening socio-economic and demographic inequality of expectations generated by the larger proportion of less advantaged individuals switching from being 'likely to unlikely' seen above.

However, the analysis so far has limitations: it cannot accommodate spells that started after age 14 (or, hence, multiple spells from one individual); and it cannot control for additional covariates. In order to relax these limitations, I now turn to multiple regression-based duration modelling techniques.

6 Multiple regression models

I estimate multiple regression duration models using the so-called 'easy estimation' methods detailed by Jenkins (1995). These are implemented using a standard binary dependent variable regression model applied to a dataset organised such that there is one observation for each time point that each individual is 'at risk' of making the transition of interest. The model exposition concentrates on the transition from 'likely to apply' to 'unlikely to apply' only to avoid unnecessary duplication; it is easy to see how the model is modified for the transition from 'unlikely to apply' to 'likely to apply'.

The outcome of interest, as outlined in Section 3, is a simple indicator of whether the individual reports being unlikely to apply to university:

$$
\begin{aligned}
Y_{it} &= 1 \text{ if young person } i \text{ is unlikely to apply to university at time } t \\
&= 0 \text{ if young person } i \text{ is likely to apply to university at time } t
\end{aligned}
\tag{2}
$$

However, as noted above, it only makes sense to include in modelling individuals who are 'at risk' of the transition in question occurring. I define a variable d_{it}, which indicates whether an individual makes the transition at a given time point, given that the individual was at risk of making the transition (i.e. they reported being likely to apply in the previous period). d_{it} takes no value where individuals are not 'at risk' of making a transition and so these observations are not included in models. The variable is formally defined as:

$$
\begin{aligned}
d_{it} &= 1 \text{ if } Y_{it} = 1 \cap Y_{it-1} = 0 \\
&= 0 \text{ if } Y_{it} = 0 \cap Y_{it-1} = 0
\end{aligned}
\tag{3}
$$

A large component of changes in young people's expectations may simply be explained by the age they have reached. If I ignore this in modelling it may result in omitted variable bias, with other covariates picking up the variation that should have been explained by age alone. I include a simple function of age in my models, denoted by α. Imposing functional form restrictions here would increase the risk of not adequately accounting for the underlying probability of transition at each age, which may also affect other estimates through omitted variable bias. Particularly because I have relatively few time periods, I use a piecewise constant age function, implemented through inclusion in the model of a dummy variable for each age (except for the first, making this the base category):

$$
\alpha(A_{it}) = \alpha_0 + \alpha_{16}A_{16.it} + \alpha_{17}A_{17.it} \tag{4}
$$

In duration models it is common to model the effect of the length of time individuals have spent in their current state on the probability of transition. A relevant example of this 'duration dependence' could be that time spent believing that you are unlikely to go to university affects one's attitudes towards and, hence, performance in school work. Such lower performance then becomes self-reinforcing of the view that you are unlikely to be in a position to apply to university. The effect of the length of time spent in a state is referred to as a 'baseline hazard rate'. In some applications, parametric 'baseline hazard functions' are used to make statements about how the underlying probability of transition changes as the length of a spell increases. However, introducing a baseline hazard function to the models in this paper has not been possible because such a large proportion of spells in the data start at the same point in time (age 14). As a result, the variables for age and time in state are highly collinear.

Since my outcome variable (d_{it}) is dichotomous, I opt to use complementary log-log regression models.[25] Using these variables and **x**, which is a vector of time-invariant and time-varying control variables (discussed further below), I estimate regression models of the form:

$$\log(-\log(1 - d_{it})) = \alpha(A_{it}) + \boldsymbol{\beta}\mathbf{x}_{it} + \varepsilon_{it} \tag{5}$$

This method of estimating duration models involves multiple observations per individual. As a consequence, ignoring the survey design, I would estimate standard errors clustered at the individual level. However, given that young people in the Longitudinal Study of Young People in England are clustered within schools, the estimated standard errors are calculated more conservatively, taking into account this higher level clustering.

I begin with a baseline model (M0), only including my age function.[26] This performs a number of roles. First, it places the survivor functions from Section 5 into this regression framework, this time allowing for multiple spells from one individual and also for spells that begin later than age 14. Second, it allows me to inspect the raw coefficients on age, providing insights on when adjustment of expectations most often occurs. Third, it provides a baseline against which I can assess the following models, in which I include additional explanatory variables.

My first model of substantive interest (M1) attempts to capture the 'total' association between SES and the probability that individuals make a transition between being 'likely' and 'unlikely' to apply. In addition to the age dummy variables, I include dummy variables indicating which quintile group of socio-economic status (SES), measured using the index described in Section 3.3, an individual is in. I leave out a variable for the third (middle) quintile group, making it the baseline category.

My second model (M2) attempts to identify the 'conditional' association between SES and the probability of making a transition, controlling for demographic characteristics, school characteristics, traumatic experiences and local labour market conditions. For demographic characteristics, the model includes gender, ethnic group, number of siblings, number of older siblings, and region of residence.

[25]The other major alternative used in duration modelling of this type are logistic models. As a robustness check, I also estimate my models using this method. Doing so makes little difference to the results.

[26]M0 does also include the month of birth and month of interview variables to try and control for the differences in age of the panel members when interviewed.

For school characteristics, I include indicators for fee-paying independent schools, selective 'grammar' schools, and for whether the school has a post-16 'sixth form'.[27] To capture the effect of traumatic experiences, I include time varying measures derived from experience of being in a workless household or having experienced a family separation. Finally, I include data proxying local labour market conditions faced by young people, specifically the local youth unemployment rate within an individual's Local Authority of residence. Since many of these variables are socially graded, I expect them to reduce the conditional association between coming from an advantaged family and the probability of transition, allowing us to assess the remaining 'effect' attributable to SES. However, as discussed in Section 3.4, the effect of SES on these variables may mean I start to underestimate the influence of SES on changes in expectations.

My third model (M3) contains the same variables as M2, and adds covariates to control for an individual's observable prior academic attainment. I include a standardised score of young people's performance at age 11 (Key Stage 2). Undoubtedly, young people's academic performance affects whether they stand a realistic chance of making a successful application to university and, hence, affects whether young people maintain their current expectations. As with some of the variables above, young people's attainment at age 11 is already likely to be affected by SES, meaning that results including prior attainment only show SES effects conditional on these results. This model is my preferred specification for identifying the 'conditional' effect of SES on changes in young people's expectations of applying to university.

My final two models specifically address whether young people's expectations are affected by the new information on their academic attainment provided by performance in examinations at age 16. The first of these (M4) adds a variable for an individual's performance in end of secondary school examinations at age 16 (Key Stage 4), standardised with mean zero and standard deviation one, and interacted with the age variable indicating that they will have received their results (age 17). As such, it will provide an estimate of the association between a one standard deviation increase in young people's performance at age 16 and the risk of transition from 'likely' to 'unlikely' or vice versa, conditional on family background and attainment at age 11. However, in interpreting this finding, it is important to note that individuals' performance in examinations at 16 is likely to be endogenous: young people's expectations of applying to university are likely to affect their effort at school and hence performance in the these examinations. As such, particular caution should be taken in the interpretation of this model. The results should only be used as indicative for the question of respon-

[27] I also estimate linear probability models including school fixed effects as a robustness check. As might be anticipated, the influence of SES is somewhat reduced in these models, but they do not alter the overall narrative.

siveness to new information on academic attainment; results from M3 are likely to be a more reliable guide to the overall association between SES and changes in young people's expectations.

The final model (M5) builds on M4, but relaxes the implicit assumption that this new information on academic performance affects all young people in the same way. I introduce an interaction between KS4 performance and SES, which allows me to explore whether individuals are more or less likely to adjust their expectations in response to their results depending on their SES background. The same caveats apply in terms of the potential endogeneity in performance at age 16, but this still provides suggestive evidence on a potentially important driver of inequality in expectations of applying to university.

Given the complexity of interpreting interaction effects, and in the interests of parsimony, I also estimate variants of models M4 and M5, in which the dummy variables for each quintile group of SES have been replaced by a single variable of my underlying SES index, standardised so that it has mean zero and standard deviation one. This simplification comes at the cost of assuming a linear relationship between my SES index and the risk of transition. However, robustness checks[28] suggest that this does not seem to affect the overall narrative of my analysis. As such, in my discussion of the results, I focus these variants, referred to as M4C and M5C.

7 Results

The results tables focus on the influence of SES on changes in expectations during this period.[29] Once again, I explore the transition from 'likely to unlikely and the transition from 'unlikely to likely' separately.

I report the results of the models using hazard ratios (exponentiated coefficients from the underlying complementary log-log regression model). These are multiplicative, rather than additive; they express no difference from the baseline group when they are equal to 1 (rather than 0, as would be the case if I were discussing coefficients). As such, when I refer to a hazard ratio being statistically significant, this means that it is statistically significantly different from 1, rather than from 0.

In models focusing on the influence of SES on transitions (M1-M3), I concentrate on the hazard ratios for each quintile group of SES, relative to a baseline category of the middle (third) quintile group. These may be interpreted as the probability that an individual in the relevant SES quintile group makes a transition, conditional on being in the state at that point, divided by the probability that an

[28]The full results of M4, M4C, M5 and M5C are reported in Appendix A for comparison.
[29]Regression tables reporting the full set of hazard ratios are reported in Appendix A.

individual in the middle SES quintile group makes a transition (conditional in the same way). In order to examine the overall patterns of young people's transitions as they age, I also report hazard ratios from each model associated with each age, relative to a baseline of the period between the interview at age 14 and age 15.

In models focusing on the responsiveness of young people to new information on their academic attainment (M4C and M5C), I concentrate on the hazard ratio associated with change in SES and the hazard ratio associated with change in both SES and KS4 performance. The former may be interpreted as the probability that an individual makes a transition, divided by the probability than an individual with one standard deviation lower SES makes a transition (conditioned as above). The latter may be interpreted as the probability that an individual makes a transition divided by the probability than an individual with one standard deviation lower SES *and* one standard deviation lower KS4 performance makes a transition.

It is also natural to want to test whether each model adds explanatory power, relative to the one before. In many circumstances this would be done with likelihood ratio tests. However, as a result of accounting for the complex survey design of the data, these are not valid. Instead, I conduct F tests of the joint significance of all additional coefficients, relative to the previous model. As the results simply show that each model does provide additional explanatory power relative to the one before, they are only reported in Appendix A.

7.1 From likely to unlikely

The results for the transition from 'likely to unlikely' are reported in Table 5. I begin by discussing the results from the baseline model (M0), to examine the point in time at which individuals currently reporting being 'likely to apply' are most likely to change to reporting being 'unlikely to apply'. The hazard ratios reported for ages 16 and 17 are statistically significantly less than one. This suggests the individuals are most likely to make a transition between their reports at age 14 and 15, with the rate of transitions slowing after this point. This reflects the Kaplan-Meier survivor function plotted in Figure 3, where the largest step was the first. However, it has commonly been observed in duration modelling that one reason for such an observation is that individuals who are most likely to make a transition have already done so before later time points (Jenkins, 2004, p.81), hence the sample at risk are systematically less likely to change their report just for this reason. Controlling for factors associated with this compositional change may, therefore, reduce the apparent effect of age.

In the first model including SES (M1), I find that the estimated hazard ratios are statistically signifi-

Table 5: Estimated hazard ratios of transition from reporting being likely to apply to reporting being unlikely to apply by quintiles of socioeconomic status

	M0	M1	M2	M3	M4
Age 16	0.89	0.90	0.91	0.95	0.94
	(-2.41)**	(-2.33)**	(-2.07)**	(-1.08)	(-1.22)
Age 17	0.74	0.77	0.85	0.92	0.92
	(-6.60)***	(-5.82)***	(-3.48)***	(-1.83)*	(-1.76)*
SES Q1 (Low)		1.46	1.54	1.13	1.10
		(6.33)***	(6.59)***	(1.80)*	(1.42)
SES Q2		1.40	1.31	1.17	1.16
		(5.61)***	(4.49)***	(2.53)**	(2.42)**
SES Q4		0.75	0.80	0.80	0.80
		(-4.76)***	(-3.69)***	(-3.67)***	(-3.71)***
SES Q5 (High)		0.33	0.39	0.47	0.47
		(-13.45)***	(-11.89)***	(-9.66)***	(-9.59)***
Significance of SES ($P > \|F\|$)		0.00	0.00	0.00	0.00
N	9,247	9,247	9,247	9,247	9,247
Variables	**M0**	**M1**	**M2**	**M3**	**M4**
Age	√	√	√	√	√
SES Quintile Dummies		√	√	√	√
Demographics & School			√	√	√
Prior Attainment				√	√
Age 16 Attainment					√

Notes: Reporting hazard ratios. $P > |F|$ shows p-value from joint significance test of the hypothesis that exponentiated coefficients on all SES group dummies in the underlying conditional log-log regression model are equal to 1. Adjusted using LSYPE-provided Wave 2 survey design and non-response weights. T-statistics of the null hypothesis that the hazard ratio is equal to one, based on standard errors clustered by individual's school, are reported in parentheses. Estimated risks are relative to base categories of Age 15 and SES quintile group 3.

cantly different from one for each of the quintile groups of SES, with young people from less advantaged backgrounds being significantly more likely to switch from reporting being 'likely' to reporting being 'unlikely'. To take the extremes, those in the least advantaged quintile group have more than four times the hazard of making a transition than those in the most advantaged SES group. In addition, the size of the change in hazard between each quintile group tends to increase further up the SES distribution: the smallest gap in hazard is between Q1 and Q2 (only equivalent to a 5 percent reduction in the probability of transition), while the largest is between Q4 and Q5 (equivalent to more than a 50% reduction in the hazard of transition). Also worthy of note is that inclusion of SES in the model has made very little difference to the correlation between age and hazard of transition.

Given previous evidence on the young people's expectations of applying to university by SES the strong relationship is unsurprising. However, the aim in the following models is to assess what, if anything, explains these gaps, and whether the SES gradient persists once other factors have been controlled for.

Moving to the second model including SES (M2), I add various demographic and school characteristics. Several of these (notably including gender, ethnicity, and school characteristics) have large hazard ratios that are statistically significantly different from one (reported in Table 9 of Appendix A). There is some reduction in the socio-economic inequalities observed in earlier models: the hazard of an individual from the least advantaged SES quintile group making a transition from 'likely to unlikely' is now estimated to be just under 4 times greater than the hazard of an individual from the most advantaged group doing so. The estimated hazard of transition for individuals in the highest SES quintile group remains dramatically different from the estimated hazard for individuals in any other quintile group: individuals have less than half the hazard of making a transition as individuals in the second most advantaged fifth of the distribution.

As anticipated, inclusion of prior academic attainment from age 11 (in M3) makes a much bigger difference to the estimated influence of SES on academic attainment. A noticeable feature of the estimated influence of SES quintile groups is that there is now no difference in the hazard of transition between the lowest two quintile groups; conditional on other characteristics, young people in the bottom 40% of the SES distribution have approximately 15% higher hazard of making a transition from 'likely to unlikely' than individuals in the middle. By contrast, the influence of being in a higher SES group continues to be large reductions in the hazard of transition from 'likely to unlikely': young people in the top SES quintile group still have approximately 50% of the hazard of making a transition as individuals in the middle.

Furthermore, introducing prior attainment reduces estimated differences in the hazard of transition by age, which become only statistically significant at a 0.1 level. This suggests that, in the case of the transition from 'likely to unlikely', much of the apparent effects of age were driven by the reduced presence in the sample of individuals with lower prior attainment by later time points.

In summary, there continues to be a strong relationship between young people's socio-economic background and their hazard of continuing to report being 'likely to apply' to university. Individuals from the least advantaged fifth of the SES distribution still have almost 2.5 times the hazard of making a transition as individuals in the most advantaged quintile group.

Table 6: Estimated odds ratios of transition from reporting being likely to apply to reporting being unlikely to apply by interaction of socio-economic status and new information on attainment at age 16

	M4C	**M5C**
Age 16	0.92	0.92
	(-1.45)	(-1.45)
Age 17	1.00	1.05
	(-0.03)	(0.84)
SES Z-Score	0.68	0.69
	(-11.00)	(-10.41)
KS4 Z-Score (After results)	0.51	0.46
	(-9.51)	(-9.98)
SES * KS4		0.79
		(-3.11)
N	9,247	9,247
Variables	**M4C**	**M5C**
Age	√	√
SES Index Z-Score	√	√
Demographics & School	√	√
Prior Attainment	√	√
Age 16 Attainment	√	√
Age 16 Attainment and SES Interaction		√

Notes: Reporting hazard ratios. Adjusted using LSYPE-provided Wave 2 survey design and non-response weights. T-statistics of the null hypothesis that the hazard ratio is equal to one, based on standard errors clustered by individual's school, are reported in parentheses. Estimated risks are relative to base category of Age 15.

What explains the reduction in the size of the SES gap once prior attainment has been included? Two possibilities are that young people from less advantaged backgrounds are less likely to have achieved strong results at age 16, for whatever reason. Alternatively, it could be that their expectations are more sensitive to the results that they receive. My final models aim to shed light on this question.

I first examine whether KS4 results do have an association with changes in young people's expectations of applying to university. I report the results from M4 in Table 5 in order to check for unexpected changes in the main effects. Given the likely endogeneity of performance at age 16, estimates from M3 are likely to be a better guide to the 'conditional' association between SES and the hazard of transition, although there are only slight changes in practice. For parsimony and ease of interpretation, at this point I switch to use of models in which SES is measured using the index variable defined in Section 3.3. Comparing the results of M4 (final column of Table 5) and M4C (first column of Table 6) suggests that this simplification does not seem to have much of an effect on other variables in the model. However, the main coefficient here is on the KS4 performance variable, which unsurprisingly shows that a one standard deviation improvement in results at age 16 are associated with a having approximately a 20% reduction in the hazard of moving from reporting 'likely to apply' to reporting 'unlikely to apply'.

Results from M5C, in the second column of Table 6, then provides evidence on the question of differing responsiveness of young people to age 16 exam results. The estimate reported in the interaction row of Table 6 should be interpreted as the additional expected change in the hazard ratio associated with a one standard deviation increase in KS4 scores when the individual in question is one standard deviation further up the SES distribution. As I do find a statistically significant estimate for this interaction term, this suggests that young people's SES background does affect how likely they are to adjust their expectations downwards when faced with a similar set of KS4 results. Specifically, the hazard ratio of 0.79 shows that, in general, young people from more advantaged backgrounds are less likely to respond to poorer results by lowering their expectations of applying to university.[30]

7.2 From unlikely to likely

I now turn to the transition back from being 'unlikely to apply' to being 'likely to apply'. I report the results in Table 7, concentrating again just on the association between young people's SES quintile group and the hazard of young people raising their expectations. As remarked above, it may well be the case that the relationship explaining the likelihood of transition from 'unlikely to likely' is quite different from that explaining 'likely to unlikely'; this could be in terms of different significant factors, different directions of effects and different strengths of relationships. However, this is not the case for the unconditional relationship between young people's age and the hazard that they make a transition from 'unlikely to likely' (in M0): as with the opposite transition, as individuals get older they appear to become less likely to switch, albeit more dramatically by age 17.

Turning to SES (in M1), once again there is a large gradient in young people's chances of making a

[30] I do also estimate separate versions of this model using dummy variables for quintiles of SES. While the results from this model suggest that a linear relationship is unlikely to provide the best fit, a joint test of the interaction terms still suggests that the overall form of the relationship reported in Table 6 is robust.

Table 7: Estimated hazard ratios of transition from reporting being unlikely to apply to reporting being likely to apply by quintiles of socioeconomic status

	M0	M1	M2	M3	M4
Age 16	0.88	0.88	0.90	0.91	0.90
	(-2.28)**	(-2.30)**	(-1.80)*	(-1.72)*	(-1.86)*
Age 17	0.63	0.63	0.63	0.64	0.76
	(-7.90)***	(-8.13)***	(-7.90)***	(-7.61)***	(-4.48)***
SES Q1 (Low)		0.76	0.70	0.79	0.81
		(-3.80)***	(-4.28)***	(-2.78)***	(-2.57)**
SES Q2		0.89	0.88	0.91	0.91
		(-1.75)*	(-1.83)*	(-1.43)	(-1.38)
SES Q4		1.29	1.25	1.16	1.15
		(3.42)***	(3.05)***	(2.00)**	(1.87)*
SES Q5 (High)		1.94	1.92	1.71	1.67
		(7.76)***	(7.68)***	(6.25)***	(5.99)***
Significance of SES ($P > \|F\|$)		0.00	0.00	0.00	0.00
N	5,330	5,330	5,330	5,330	5,330
Variables	**M0**	**M1**	**M2**	**M3**	**M4**
Age	√	√	√	√	√
SES Quintile Dummies		√	√	√	√
Demographics & School			√	√	√
Prior Attainment				√	√
Age 16 Attainment					√

Notes: Reporting hazard ratios. $P > |F|$ shows p-value from joint significance test of the hypothesis that exponentiated coefficients on all SES group dummies in the underlying conditional log-log regression model are equal to 1. Adjusted using LSYPE-provided Wave 2 survey design and non-response weights. T-statistics of the null hypothesis that the hazard ratio is equal to one, based on standard errors clustered by individual's school, are reported in parentheses. Estimated risks are relative to base categories of Age 15 and SES quintile group 3.

transition depending on their relative advantage. In this case, young people from more advantaged backgrounds have a greater hazard of making a transition from reporting ‘unlikely’ to reporting ‘likely’. Individuals from the most advantaged quintile group of the SES index have more than 2.5 times the hazard of making a transition as their counterparts in the least advantaged fifth of the distribution. This is a large difference, although not as large as the difference between these groups in the hazard of moving from ‘likely to unlikely’, where the unconditional hazard ratio was greater than four. However, as with the inverse transition, will this apparent influence of SES be reduced when I add further covariates?

The additional covariates in M2 do nothing to reduce the association between SES and the hazard of making a transition from ‘unlikely to likely’. The hazard ratios barely change for any of the quintile groups of SES. Coefficients on some of the variables added at this point (reported in Table 10 of Appendix A) suggest large and significant relationships with the hazard of transition: in particular young people who from ethnic minorities and young women are much more likely to switch to being ‘likely to apply’. However, the results suggest that these are largely independent of SES and/or cancel one another out.

On the other hand, controlling for prior attainment does more to explain the SES influence on young people’s chances of changing their minds from ‘unlikely to likely’, particularly at the more advantaged end of the SES distribution. Nevertheless, a large SES gradient remains, with individuals in the top quintile group of the SES index having more than twice the hazard of moving from ‘unlikely’ to ‘likely’ as peers in the bottom group. The most advantaged fifth of the sample remain outliers from the rest of the distribution: their hazard of transition is almost fifty percent higher than in the quintile group just below them.

In contrast to the results for ‘likely to unlikely’, the coefficients on whether an individual attends an independent school, a grammar school, or a school with a sixth form are not statistically significant. However, it would appear that in the former two cases this is due to there only being a very small number of such individuals in the sample on which models of the transition from ‘unlikely to likely’ are estimated: there are very few individuals from independent or grammar schools who ever report being ‘unlikely to apply’ to university during this period.

Another noticeable difference between the two directions of transition is that, in contrast to the model of ‘likely to unlikely’, even inclusion of young people’s prior attainment in the model of ‘unlikely to likely’ does not fully explain the role of age: the coefficient on age 16 becomes only significant at the 10% level, while the coefficient on age 17 remains highly significant. One explanation for this is that, while it's never too late to decide against making an application to university, it can get too late for individuals to start thinking that they will. If they have not been planning to apply to university, young people will not have taken key actions necessary in order to be in a position to make a competitive application. Arguably this is closer to a duration effect than an age effect, being picked up by the age variables due to the absence of duration parameters: it is less likely to be present for young people who only spend a single period reporting being 'unlikely to apply', for example.

In summary, as with the transition from 'likely to unlikely', there remains a large, statistically significant relationship between young people's socio-economic advantage and the likelihood that they move into thinking they are 'likely to apply'.

Table 8: Estimated odds ratios of transition from reporting being unlikely to apply to reporting being likely to apply by interaction of socio-economic status and new information on attainment at age 16

	M4C	**M5C**
Age 16	0.88	0.88
	(-1.92)	(-1.94)
Age 17	0.75	0.73
	(-4.02)	(-4.29)
SES Z-Score	1.34	1.35
	(7.20)	(7.34)
KS4 Z-Score (After results)	1.84	2.06
	(8.32)	(8.05)
SES * KS4		1.22
		(2.32)
N	5,330	5,330
Variables	**M4C**	**M5C**
Age	√	√
SES Index Z-Score	√	√
Demographics & School	√	√
Prior Attainment	√	√
Age 16 Attainment	√	√
Age 16 Attainment and SES Interaction		√

Notes: Reporting hazard ratios. Adjusted using LSYPE-provided Wave 2 survey design and non-response weights. T-statistics of the null hypothesis that the hazard ratio is equal to one, based on standard errors clustered by individual's school, are reported in parentheses. Estimated risks are relative to base category of Age 15.

Again, the question arises of whether young people from less advantaged backgrounds are responding differently to new information on their academic attainment. Specifically, in this case, the hypothesis that may partially explain the growing inequality in expectations is that individuals from lower SES backgrounds are less responsive to just as promising new information at age 16 as peers with similar prior academic attainment from more advantaged homes. As with the transition from 'likely to unlikely', I switch at this point to use of a continuous measure of SES. As such, in Table 8, the estimate reported in the interaction row (SES * KS4) reports the additional expected change in the risk of transition associated with a one standard deviation increase in KS4 scores when the individual in question is one standard deviation further up the SES distribution.

Indeed, the results do suggest differential sensitivity to new information on academic performance may be important in explaining the observed changes in expectations. There is a statistically significant hazard ratio of 1.29 associated with the interaction term,[31] suggesting that individuals with the same age 16 performance but with more advantaged parents are more likely to revise their expectations in light of better academic results at age 16.

8 Conclusions

In this paper, I have investigated how young people's expectations of applying to university change between age 14 and age 17, just before individuals start making applications. My findings confirm that this is a period when many young people do change their expectations of applying to university. They also highlight that this change is not just from being 'likely to apply' to being 'unlikely to apply', but rather runs in both directions.

While young people across the socio-economic status distribution start their adolescence with high educational expectations, those from less advantaged backgrounds are much more likely to revise their expectations downwards and much less likely to raise their expectations during this period. This relationship persists even once I control for many other factors correlated with SES and, perhaps most notably, young people's prior academic attainment. The least advantaged fifth of young people have more than twice the chances of switching from reporting being 'likely to apply' to reporting being 'unlikely to apply' as the most advantaged fifth, conditional on prior attainment. Conversely, the most advantaged fifth of young people have more than twice the chances of changing from reporting being 'unlikely to apply' to reporting being 'likely to apply' as the most advantaged fifth, again conditional on prior attainment.

In previous work I found that much of the socio-economic gradient in access to university opened at or before the point of application (Anders, 2012a). This paper builds on this, finding that a substantial portion of this socio-economic gap in university applications opens between ages 14 and 17. A positive implication of this is that it is not too late to target policies, both to maintain and to raise educational expectations, at bright individuals from less advantaged backgrounds during this period of their lives. However, of the two, raising expectations of applying to university may be less effective

[31]As with the model from 'likely to unlikely', the results from a separate model model where I use dummy variables for quintile groups of SES suggest that a linear relationship is unlikely to provide the best fit. Nevertheless, in a model in which dummy variables are used, a joint test of the interaction terms suggests this finding is robust.

than maintaining expectations and becomes increasingly difficult as individuals get older.

I also find some evidence that young people from differing SES backgrounds react differently to new information on their academic attainment at age 16. This differential is also asymmetric, helping to explain the growth in inequality of expectations: more advantaged young people are less responsive to results in lowering their expectations, but more responsive to results in raising them. After these exam results is a difficult point in time to reach young people, as many move between educational institutions or leave full time education altogether. However, it may be the case that providing fresh guidance in the light of the results is very important in ensuring young people's educational expectations are appropriate.

Bibliography

Alcott, B. (2013a). Predicting departure from British education: Identifying those most at risk through discrete time hazard modelling. *Widening Participation and Lifelong Learning*, 15(4):46–64. doi:10.5456/WPLL.15.4.46.

Alcott, B. (2013b). The Influence of Teacher Encouragement on Educational Persistence: Evidence From England. Working Paper, School of Education, University of Michigan.

Anders, J. (2012a). The link between household income, university applications and university attendance. *Fiscal Studies*, 33(2):185–210. doi:10.1111/j.1475-5890.2012.00158.x.

Anders, J. (2012b). Using the Longitudinal Study of Young People in England (LSYPE)/"Next Steps" for research into Higher Education access. DoQSS Working Paper 12-13, Institute of Education, University of London. Available from http://repec.ioe.ac.uk/repec/pdf/qsswp1213.pdf.

Anders, J. and Micklewright, J. (2013). Teenagers' expectations of applying to university: how do they change? DoQSS Working Paper 13, Department of Quantiative Social Science, Institute of Education, University of London. Available from http://repec.ioe.ac.uk/REPEc/pdf/qsswp1313.pdf.

Bane, M. J. and Ellwood, D. T. (1986). Slipping into and out of poverty: the dynamics of spells. *The Journal of Human Resources*, 21(4):1–23. Available from http://www.jstor.org/stable/145955.

BBC News. Tests scrapped for 14-year-olds. Retrieved from http://news.bbc.co.uk/1/hi/education/7669254.stm on 28/09/2014.

Box-Steffensmeier, J. M. and Jones, B. S. (2004). *Event History Modeling: A Guide for Social Scientists*. Analytical Methods for Social Research. Cambridge University Press, Cambridge.

Breen, R. and Yaish, M. (2006). Testing the Breen-Goldthorpe Model of Educational Decision-Making. In Morgan, S. L., Grusky, D. B., and Fields, G. S., editors, *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, pages 232–258. Stanford University Press, Redwood City.

Cameron, S. V. and Heckman, J. J. (1999). Can tuition policy combat rising wage inequality? In Kosters, M. H., editor, *Financing College Tuition: Government Policies and Educational Priorities*, chapter 5, pages 76–124. The AEI Press, Washington, D.C.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A*, 176(2):431–457. doi:10.1111/j.1467-985X.2012.01043.x.

Chowdry, H., Crawford, C., and Goodman, A. (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16. *Longitudinal and Life Course Studies*, 2(1):59–76. doi:10.14301/llcs.v2i1.141.

Clark, T. G., Bradbury, M. J., Love, S. B., and Altman, D. G. (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, 89:232–238. doi:10.1038/sj.bjc.6601118.

Collingwood, A., Cheshire, H., Nicolaas, G., D'Souza, J., Ross, A., Hall, J., Armstrong, C., Prosser, A., Green, R., Collins, D., Gray, M., and McNaughton Nicholls, C. (2010). A review of the Longitudinal Study of Young People in England (LSYPE): recommendations for a second cohort. DfE Research Report DFE-RR048, Department for Education. Available from https://www.gov.uk/government/publications/a-review-of-the-longitudinal-study-of-young-people-in-england-lsype-recommendations-for-a-second-cohort.

DesJardins, S. L. (2003). Event History Methods: Conceptual Issues and an Application to Student Departure from College. In *Higher Education: Handbook of Theory and Research*, volume 18, chapter 8, pages 421–471. Kluwer Academic Publishers.

Fumagalli, L. (2012). Great expectations. Channels and barriers to university education. ISER Working paper, Institute of Social and Economic Research, University of Essex. Available from https://www.iser.essex.ac.uk/publication/521127.

Ganzach, Y. (2000). Parents' education, cognitive ability, educational expectations and educational attainment: Interactive effects. *British Journal of Educational Psychology*, 70(3):419–441. doi:10.1348/000709900158218.

Goldthorpe, J. H. and McKnight, A. (2004). The economic basis of social class. CASEpaper 80, Centre for Analysis of Social Exclusion, London School of Economics. Available from http://sticerd.lse.ac.uk/dps/case/cp/CASEpaper80.pdf.

Goodliffe, J. (2003). The hazards of time-varying covariates. Working paper, Department of Political Science, Brigham Young University. Available from http://goodliffe.byu.edu/papers/tvc2.pdf.

Gorard, S. (2012). Querying the causal role of attitudes in educational attainment. *ISRN Education*, 2012:Article ID 501589. doi:10.5402/2012/501589.

Gottfredson, L. S. (2002). Gottfredson's theory of circumscription, compromise, and self-creation. In Brown, D., editor, *Career Choice and Development*. Jossey-Bass, San Fransisco, CA.

Gutman, L. M. and Akerman, R. (2008). Determinants of aspirations. Research Report 27, Centre for Research on the Wider Benefits of Learning, Institute of Education, University of London. Available from http://eprints.ioe.ac.uk/2052/1/Gutman2008Determinants.pdf.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, 65(6):1541–1545. doi:10.2307/1131279.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320. doi:10.2307/1911491.

Iceland, J. (1997). The dynamics of povery spells and issues of left-censoring. PSC Research Report 97-378, Populations Studies Center, University of Michigan. Available from http://www.psc.isr.umich.edu/pubs/pdf/rr97-378.pdf.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1):129–138. doi:10.1111/j.1468-0084.1995.tb00031.x.

Jenkins, S. P. (2004). Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex. Available from http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf.

Jenkins, S. P. and Schluter, C. (2002). The effect of family income during childhood on later-life attainment: Evidence from germany. IZA Discussion Paper 604, Institute for the Study of Labor. Available from http://ssrn.com/paper=343340.

Jerrim, J. (2011). Disadvantaged children's "low" educational expectations: Are the US and UK really so different to other industrialized nations? DoQSS Working Paper 11-04, Department of Quantitative Social Science, Institute of Education. Available from http://repec.ioe.ac.uk/repec/pdf/qsswp1104.pdf.

Kao, G. and Tienda, M. (1998). Educational aspirations of minority youth. *American Journal of Education*, 106(3):349–384. Available from http://www.jstor.org/stable/1085583.

Khoo, S. T. and Ainley, J. (2005). Attitudes, intentions and participation. Longitudinal Surveys of Australian Youth Research Report 41, Australian Council for Educational Research. Available from http://www.lsay.edu.au/publications/1847.html.

Kolenikov, S. and Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*, 55(1):128–165. doi:10.1111/j.1475-4991.2008.00309.x.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782. doi:10.2307/2938349.

Morgan, S. L. (1998). Adolescent educational expectations: Rationalized, fantasized, or both? *Rationality and Society*, 10(2):131–162. doi:10.1177/104346398010002001.

Office for National Statistics (2004). Annual Population Survey. Accessed via nomis (https://www.nomisweb.co.uk).

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460. doi:10.1007/BF02296207.

Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4):805–827. doi:10.1111/j.1467-985X.2006.00426.x.

Rampino, T. and Taylor, M. (2013). Gender differences in educational aspirations and attitudes. ISER Working Paper 2013-15, Institute for Social and Economic Research, University of Essex. Available from https://www.iser.essex.ac.uk/publications/working-papers/iser/2013-15.pdf.

Reynolds, J. R. and Pemberton, J. (2001). Rising College Expectations among Youth in the United States: A Comparison of the 1979 and 1997 NLSY. *The Journal of Human Resources*, 36(4):703–726. Available from http://www.jstor.org/stable/3069639.

Rose, D. and Pevalin, D. J. (2001). The National Statistics Socio-economic Classification: Unifying Official and Sociological Approaches to the Conceptualisation and Measurement of Social Class. ISER Working Paper 2001-4, Institute for Social and Economic Research, University of Essex. Available from https://www.iser.essex.ac.uk/publications/working-papers/iser/2001-04.pdf.

Rose, D., Pevalin, D. J., and O'Reilly, K. (2005). *The National Statistics Socio-economic Classification: Origins, Development and Use*. Palgrave Macmillan/Office for National Statistics, Basingstoke, UK.

Sanders, M., Kristal, A., Sabri, F., and Tupper, A. (2013). Aspiration and inspiration - a pilot study of mentoring in schools. CMPO Working Paper 13/314, Centre for Market and Public Organisation, University of Bristol. Available from http://www.bristol.ac.uk/cmpo/publications/papers/2013/wp314.pdf.

Schoon, I., Barnes, M., Brown, V., Parsons, S., Ross, A., and Vignoles, A. (2012). Intergenerational transmission of worklessness: Evidence from the Millennium Cohort and Longitudinal Study of Young People in England. DfE Research Report DFE-RR234, Department for Education. Available from https://www.gov.uk/government/publications/intergenerational-transmission-of-worklessness-evidence-from-the-millennium-cohort-and-the-longitudinal-study-of-young-people-in-england.

StataCorp (2013). *Stata Survival Analysis and Epidemiological Tables Reference Manual: Release 13*. Stata Press, College Station, TX.

Steele, F. (2005). Event history analysis. NCRM Methods Review Paper NCRM/004, ESRC National Centre for Research Methods. Available from http://eprints.ncrm.ac.uk/88/1/MethodsReviewPaperNCRM-004.pdf.

Strand, S. (2014). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education*, 40(2):223–245. doi:10.1080/03054985.2014.891980.

Thornton, A., Pickering, E., Peters, M., Leathwood, C., Hollingworth, S., and Mansaray, A. (2014). School and college-level strategies to raise aspirations of high-achieving disadvantaged pupils to pursue higher education investigation. DfE Research Report DFE-RR296, Department for Education. Available from https://www.gov.uk/government/publications/school-level-strategies-to-raise-aspirations-to-higher-education.

UCAS (2012). End of cycle report 2012. Report, UCAS.

van den Berg, G. J., Lindenboom, M., and Dolton, P. J. (2006). Survey non-response and the duration of unemployment. *Journal of the Royal Statistical Society. Series A (Statistics in Society*, 169(3):585–604. Available from http://www.jstor.org/stable/3877437.

Vermunt, J. K. (2001). Event history analysis, unobserved heterogeneity. Working paper, Department of Methodology, Faculty of Social and Social and Behavioral Sciences, Tilburg University. Available from http://arno.uvt.nl/show.cgi?fid=13811.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, London.

A Full regression tables

Table 9: Estimated effects on risk of transition from reporting being 'likely to apply' to university to reporting being 'unlikely to apply' to university: hazard ratios

	M0	M1	M2	M3	M4	M4C	M5	M5C
Age 16	0.89	0.90	0.91	0.95	0.94	0.95	0.95	0.95
	(0.04)**	(0.04)**	(0.04)**	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Age 17	0.74	0.77	0.85	0.92	0.92	0.92	0.97	0.97
	(0.03)***	(0.03)***	(0.04)***	(0.04)*	(0.05)*	(0.05)*	(0.05)	(0.05)
SES Q1 (Low)		1.46	1.54	1.13	1.10		1.14	
		(0.09)***	(0.10)***	(0.08)*	(0.07)		(0.08)**	
SES Q2		1.40	1.31	1.17	1.16		1.16	
		(0.08)***	(0.08)***	(0.07)**	(0.07)**		(0.07)**	
SES Q4		0.75	0.80	0.80	0.80		0.80	
		(0.05)***	(0.05)***	(0.05)***	(0.05)***		(0.05)***	
SES Q5 (High)		0.33	0.39	0.47	0.47		0.49	
		(0.03)***	(0.03)***	(0.04)***	(0.04)***		(0.04)***	
SES Z-Score						0.71		0.72
						(0.02)***		(0.02)***
Male			1.49	1.53	1.49	1.50	1.50	1.51
			(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Ethnicity: Mixed			0.63	0.63	0.62	0.62	0.61	0.60
			(0.06)***	(0.07)***	(0.07)***	(0.06)***	(0.07)***	(0.06)***
Ethnicity: Indian			0.19	0.17	0.17	0.16	0.17	0.16
			(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Ethnicity: Pakistani			0.27	0.23	0.23	0.22	0.23	0.22
			(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Bangladeshi			0.26	0.27	0.27	0.25	0.27	0.25
			(0.05)***	(0.05)***	(0.05)***	(0.04)***	(0.05)***	(0.04)***
Ethnicity: Black Caribbean			0.37	0.27	0.26	0.27	0.26	0.27
			(0.05)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Ethnicity: Black African			0.20	0.17	0.17	0.17	0.17	0.17
			(0.04)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Other			0.27	0.24	0.25	0.24	0.24	0.23
			(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
Attended Independent School			0.30	0.27	0.29	0.29	0.29	0.30
			(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Attended Grammar School			0.23	0.37	0.38	0.37	0.39	0.38
			(0.05)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Attended School with Sixth Form			0.84	0.86	0.86	0.87	0.86	0.86
			(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Experienced workless household			0.97	0.88	0.83	0.76	0.85	0.78
			(0.06)	(0.06)**	(0.05)***	(0.05)***	(0.05)***	(0.05)***
Ever experienced family separation			0.96	0.95	0.94	0.94	0.94	0.95
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Local Youth Unemployment Rate / 10			0.96	0.95	0.95	0.95	0.96	0.95
			(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
KS2 Z-Score				0.56	0.60	0.60	0.60	0.61
				(0.01)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
KS4 Z-Score (After results)					0.67	0.67	0.50	0.60
					(0.03)***	(0.03)***	(0.06)***	(0.03)***
KS4 Z-Score * SES Q1							1.62	
							(0.25)***	
KS4 Z-Score * SES Q2							1.47	
							(0.22)**	
KS4 Z-Score * SES Q4							1.19	
							(0.18)	
KS4 Z-Score * SES Q5							0.96	
							(0.21)	
KS4 Z-Score * SES Z-Score								0.79
								(0.05)***
Geographical			√	√	√	√	√	√
Number and order of siblings			√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√	√
F test of difference from previous model	.	113.10	25.82	248.18	63.78	101.97	3.77	16.26
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Number of individuals	9,247	9,247	9,247	9,247	9,247	9,247	9,247	9,247

Notes: Reporting hazard ratios. Standard errors (clustered by individual's school) in parentheses. Weighted using Wave 2 survey design and non-response weights. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.

Table 10: Estimated effects on risk of transition from reporting being 'unlikely to apply' to university to reporting being 'likely to apply' to university: hazard ratios

	M0	M1	M2	M3	M4	M4C	M5	M5C
Age 16	0.88	0.88	0.90	0.91	0.90	0.90	0.90	0.90
	(0.05)**	(0.05)**	(0.05)*	(0.05)*	(0.05)*	(0.05)*	(0.05)*	(0.05)*
Age 17	0.63	0.63	0.63	0.64	0.76	0.76	0.75	0.74
	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
SES Q1 (Low)		0.76	0.70	0.79	0.81		0.79	
		(0.06)***	(0.06)***	(0.07)***	(0.07)**		(0.07)***	
SES Q2		0.89	0.88	0.91	0.91		0.91	
		(0.06)*	(0.06)*	(0.06)	(0.06)		(0.06)	
SES Q4		1.29	1.25	1.16	1.15		1.13	
		(0.10)***	(0.09)***	(0.08)**	(0.08)*		(0.08)*	
SES Q5 (High)		1.94	1.92	1.71	1.67		1.68	
		(0.17)***	(0.16)***	(0.15)***	(0.14)***		(0.14)***	
SES Z-Score						1.28		1.29
						(0.04)***		(0.04)***
Male			0.60	0.60	0.62	0.62	0.62	0.62
			(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Mixed			1.50	1.55	1.56	1.54	1.58	1.55
			(0.19)***	(0.20)***	(0.20)***	(0.20)***	(0.20)***	(0.20)***
Ethnicity: Indian			2.85	3.33	3.23	3.27	3.24	3.27
			(0.48)***	(0.51)***	(0.50)***	(0.51)***	(0.51)***	(0.50)***
Ethnicity: Pakistani			3.62	4.27	4.17	4.35	4.18	4.31
			(0.44)***	(0.58)***	(0.55)***	(0.58)***	(0.55)***	(0.57)***
Ethnicity: Bangladeshi			4.69	5.26	4.92	5.16	4.96	5.17
			(0.61)***	(0.70)***	(0.65)***	(0.67)***	(0.66)***	(0.67)***
Ethnicity: Black Caribbean			2.77	3.21	3.15	3.08	3.20	3.10
			(0.43)***	(0.47)***	(0.45)***	(0.45)***	(0.46)***	(0.45)***
Ethnicity: Black African			4.87	6.40	6.08	6.15	6.11	6.11
			(1.01)***	(1.35)***	(1.27)***	(1.35)***	(1.28)***	(1.33)***
Ethnicity: Other			3.15	3.56	3.53	3.64	3.53	3.62
			(0.49)***	(0.62)***	(0.59)***	(0.59)***	(0.60)***	(0.59)***
Attended Independent School			1.29	1.37	1.33	1.34	1.32	1.32
			(0.39)	(0.36)	(0.33)	(0.34)	(0.33)	(0.33)
Attended Grammar School			1.77	1.05	0.99	0.96	0.96	0.94
			(0.34)***	(0.22)	(0.20)	(0.20)	(0.20)	(0.19)
Attended School with Sixth Form			1.07	1.04	1.04	1.03	1.04	1.03
			(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Experienced workless household			0.99	1.03	1.07	1.09	1.07	1.08
			(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)
Ever experienced family separation			1.09	1.09	1.09	1.09	1.10	1.10
			(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Local Youth Unemployment Rate / 10			1.06	1.03	1.04	1.03	1.04	1.03
			(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
KS2 Z-Score				1.55	1.45	1.45	1.45	1.45
				(0.05)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
KS4 Z-Score (After results)					1.73	1.74	1.90	1.88
					(0.11)***	(0.12)***	(0.31)***	(0.14)***
KS4 Z-Score * SES Q1							0.80	
							(0.14)	
KS4 Z-Score * SES Q2							0.91	
							(0.19)	
KS4 Z-Score * SES Q4							1.43	
							(0.36)	
KS4 Z-Score * SES Q5							0.67	
							(0.16)*	
KS4 Z-Score * SES Z-Score								1.18
								(0.09)**
Geographical			√	√	√	√	√	√
Number and order of siblings			√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√	√
F test of difference from previous model	.	34.70	14.62	110.58	69.98	68.80	2.50	4.54
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.00	0.04	0.03
Number of individuals	5,330	5,330	5,330	5,330	5,330	5,330	5,330	5,330

Notes: Reporting hazard ratios. Standard errors (clustered by individual's school) in parentheses. Weighted using Wave 2 survey design and non-response weights. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.

B Weighting data using final wave attrition weights

One of the advantages of duration modelling is that we can treat missing outcome data at ‘censored’, rather than having top drop the respondent from our analysis. However, doing so will only result in unbiased estimates under the assumption that missing data censoring is ‘uninformative’ (Clark et al., 2003, p.236). In this appendix, I repeat my analysis, restricting the sample only to those still participating in the survey at Wave 4 (when the response rate relative to Wave 1 has fallen to 73% (Collingwood et al., 2010, p.52)), and weighting the analysis the LSYPE-provided attrition and non-response weights for Wave 4.

In other respects, the regression setup remains the same as for the analysis in the main body of the paper. I report the results from these analyses in Tables 11 and 12. Reassuringly, I do not find any qualitative differences from the results presented in the main paper.

C Multiple regression models accounting for unobserved heterogeneity

Unobserved heterogeneity is a problem in many statistical analyses. However, it has the potential to cause particular bias in the case of duration analysis, including “downward bias in the time effects [and, as a result,] spurious effects of time-varying covariates” (Vermunt, 2001, p.1). These are caused by changes in the composition of the sample we are analysing at each time point: individuals who are still at risk at later time points are less likely to switch to reporting being ‘unlikely to apply’ partly because the most likely to switch have already done so. Obviously, some of the characteristics in the model will control for observable changes in composition, but not all of such changes will be observable. In addition, attempting to account for unobserved heterogeneity also helps to account for the shared covariance of using multiple spells from the same individual (Steele, 2005, p.16-19).

Many duration models attempt to control for unobserved heterogeneity between individuals.[32] A popular method to account for unobserved heterogeneity is by introducing an individual-level random effect (Wooldridge, 2002, ch.10). These still allow inclusion of individual-level (i.e. non-time-varying) covariates and are relatively efficient, which is important when there are only a small number of observations for each individual. However, it makes the assumption that the individual-level random effect is not correlated with the included explanatory variables, which is almost certainly not strictly justified.

The alternative that does not make this assumption (nor any assumption about the distribution of

[32]These are often referred to as ‘frailty’ models, since, in epidemiological applications, the unobserved propensity of an individual to fall sick could be thought of as their frailty.

Table 11: Estimated effects on risk of transition from reporting being ‘likely to apply’ to university to reporting being ‘unlikely to apply’ to university: hazard ratios (Wave 4 weights applied, excludes individuals not in sample at age 17)

	M0	M1	M2	M3	M4	M4C	M5	M5C
Age 16	0.90	0.90	0.91	0.96	0.96	0.96	0.96	0.96
	(0.04)**	(0.04)**	(0.04)*	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Age 17	0.80	0.84	0.93	1.01	1.02	1.03	1.07	1.08
	(0.04)***	(0.04)***	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
SES Q1 (Low)		1.51	1.57	1.15	1.12		1.14	
		(0.09)***	(0.11)***	(0.08)**	(0.08)*		(0.08)*	
SES Q2		1.43	1.32	1.17	1.15		1.15	
		(0.09)***	(0.08)***	(0.07)**	(0.07)**		(0.07)**	
SES Q4		0.74	0.79	0.79	0.79		0.78	
		(0.05)***	(0.05)***	(0.05)***	(0.05)***		(0.05)***	
SES Q5 (High)		0.33	0.39	0.47	0.47		0.48	
		(0.03)***	(0.03)***	(0.04)***	(0.04)***		(0.04)***	
SES Z-Score						0.72		0.72
						(0.02)***		(0.02)***
Male			1.46	1.49	1.47	1.47	1.47	1.48
			(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Ethnicity: Mixed			0.65	0.63	0.61	0.62	0.60	0.59
			(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Ethnicity: Indian			0.18	0.16	0.17	0.16	0.16	0.16
			(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Ethnicity: Pakistani			0.26	0.21	0.22	0.20	0.22	0.20
			(0.03)***	(0.03)***	(0.03)***	(0.02)***	(0.03)***	(0.02)***
Ethnicity: Bangladeshi			0.26	0.26	0.27	0.25	0.27	0.25
			(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
Ethnicity: Black Caribbean			0.37	0.28	0.27	0.27	0.26	0.27
			(0.05)***	(0.05)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Ethnicity: Black African			0.17	0.15	0.15	0.15	0.14	0.15
			(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Other			0.27	0.24	0.24	0.24	0.24	0.23
			(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
Attended Independent School			0.27	0.24	0.26	0.27	0.27	0.28
			(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***
Attended Grammar School			0.24	0.39	0.39	0.39	0.40	0.40
			(0.05)***	(0.08)***	(0.08)***	(0.08)***	(0.08)***	(0.08)***
Attended School with Sixth Form			0.84	0.85	0.85	0.85	0.85	0.85
			(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Experienced workless household			1.02	0.95	0.89	0.82	0.91	0.85
			(0.06)	(0.06)	(0.06)*	(0.06)***	(0.06)	(0.06)**
Ever experienced family separation			0.98	0.97	0.93	0.93	0.95	0.95
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Local Youth Unemployment Rate / 10			0.95	0.93	0.94	0.93	0.94	0.93
			(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
KS2 Z-Score				0.56	0.60	0.61	0.61	0.61
				(0.01)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
KS4 Z-Score (After results)					0.61	0.62	0.44	0.56
					(0.03)***	(0.03)***	(0.05)***	(0.03)***
KS4 Z-Score * SES Q1							1.56	
							(0.22)***	
KS4 Z-Score * SES Q2							1.60	
							(0.23)***	
KS4 Z-Score * SES Q4							1.27	
							(0.19)	
KS4 Z-Score * SES Q5							1.06	
							(0.22)	
KS4 Z-Score * SES Z-Score								0.82
								(0.05)***
Geographical			√	√	√	√	√	√
Number and order of siblings			√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√	√
F test of difference from previous model	.	118.90	25.78	258.97	90.29	110.35	3.86	11.16
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Number of individuals	8,616	8,616	8,616	8,616	8,616	8,616	8,616	8,616

Notes: Reporting hazard ratios. Standard errors (clustered by individual’s school) in parentheses. Weighted using Wave 4 survey design and non-response weights. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.

Table 12: Estimated effects on risk of transition from reporting being 'unlikely to apply' to university to reporting being 'likely to apply' to university: hazard ratios (Wave 4 weights applied, excludes individuals not in sample at age 17)

	M0	M1	M2	M3	M4	M4C	M5	M5C
Age 16	0.88	0.88	0.90	0.91	0.90	0.90	0.90	0.90
	(0.05)**	(0.05)**	(0.05)*	(0.05)	(0.05)*	(0.05)*	(0.05)*	(0.05)*
Age 17	0.69	0.68	0.68	0.69	0.81	0.81	0.79	0.79
	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
SES Q1 (Low)		0.77	0.70	0.81	0.82		0.80	
		(0.06)***	(0.06)***	(0.07)***	(0.07)**		(0.07)***	
SES Q2		0.90	0.89	0.92	0.93		0.92	
		(0.06)	(0.06)	(0.06)	(0.06)		(0.06)	
SES Q4		1.34	1.28	1.19	1.18		1.17	
		(0.10)***	(0.10)***	(0.09)**	(0.09)**		(0.09)**	
SES Q5 (High)		1.98	1.97	1.73	1.69		1.71	
		(0.17)***	(0.17)***	(0.15)***	(0.15)***		(0.15)***	
SES Z-Score						1.29		1.29
						(0.04)***		(0.05)***
Male			0.60	0.60	0.61	0.62	0.61	0.62
			(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Mixed			1.50	1.57	1.58	1.56	1.60	1.57
			(0.22)***	(0.22)***	(0.22)***	(0.22)***	(0.22)***	(0.22)***
Ethnicity: Indian			2.73	3.26	3.17	3.20	3.17	3.19
			(0.47)***	(0.52)***	(0.52)***	(0.52)***	(0.51)***	(0.51)***
Ethnicity: Pakistani			3.66	4.24	4.15	4.33	4.15	4.29
			(0.49)***	(0.63)***	(0.60)***	(0.62)***	(0.60)***	(0.62)***
Ethnicity: Bangladeshi			5.02	5.71	5.31	5.58	5.33	5.58
			(0.69)***	(0.79)***	(0.72)***	(0.75)***	(0.72)***	(0.75)***
Ethnicity: Black Caribbean			2.72	3.12	3.08	3.01	3.14	3.03
			(0.46)***	(0.51)***	(0.49)***	(0.48)***	(0.50)***	(0.48)***
Ethnicity: Black African			5.83	8.68	7.90	8.07	7.99	8.10
			(1.20)***	(1.72)***	(1.57)***	(1.68)***	(1.61)***	(1.69)***
Ethnicity: Other			3.32	3.73	3.72	3.86	3.69	3.80
			(0.57)***	(0.71)***	(0.67)***	(0.68)***	(0.67)***	(0.68)***
Attended Independent School			1.32	1.41	1.35	1.36	1.34	1.34
			(0.39)	(0.35)	(0.33)	(0.34)	(0.33)	(0.33)
Attended Grammar School			1.75	1.00	0.95	0.92	0.92	0.90
			(0.36)***	(0.22)	(0.21)	(0.20)	(0.20)	(0.20)
Attended School with Sixth Form			1.08	1.05	1.04	1.04	1.05	1.04
			(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Experienced workless household			1.04	1.09	1.13	1.15	1.13	1.14
			(0.08)	(0.08)	(0.09)	(0.08)*	(0.09)	(0.08)*
Ever experienced family separation			1.13	1.14	1.14	1.14	1.15	1.14
			(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Local Youth Unemployment Rate / 10			1.08	1.05	1.05	1.05	1.05	1.05
			(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
KS2 Z-Score				1.58	1.49	1.48	1.48	1.48
				(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
KS4 Z-Score (After results)					1.64	1.65	1.74	1.78
					(0.11)***	(0.11)***	(0.31)***	(0.14)***
KS4 Z-Score * SES Q1							0.84	
							(0.16)	
KS4 Z-Score * SES Q2							0.96	
							(0.22)	
KS4 Z-Score * SES Q4							1.52	
							(0.40)	
KS4 Z-Score * SES Q5							0.70	
							(0.18)	
KS4 Z-Score * SES Z-Score								1.17
								(0.09)**
Geographical			√	√	√	√	√	√
Number and order of siblings			√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√	√
F test of difference from previous model	.	33.73	13.10	114.97	53.38	60.11	2.44	4.08
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.00	0.05	0.04
Number of individuals	4,864	4,864	4,864	4,864	4,864	4,864	4,864	4,864

Notes: Reporting hazard ratios. Standard errors (clustered by individual's school) in parentheses. Weighted using Wave 4 survey design and non-response weights. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.

the unobserved heterogeneity) is estimation of individual-level fixed effects. However, this approach would prevent me from being able to estimate the effect of any time-invariant covariates, which are matters of interest for this paper. Furthermore, it is unlikely that the individual-level fixed effect would be well estimated with so few observations per person in many cases: this can cause its own problems (Vermunt, 2001, p.11-12). As such, despite its assumptions not being fully met, I use random effects modelling. This is preferable to simply assuming unobserved heterogeneity is not an issue.

One must also make an assumption about the distribution of the individual-level random effects, with popular distributions including the Gamma distribution (Meyer, 1990), a normal distribution with mean zero (Jenkins, 2004, ch. 8.2), or non-parametric discrete mixing distribution (latent class analysis) (Heckman and Singer, 1984). For the models reported in this section, I assume a normal distribution for the random effects. However, I have also estimated models with a discrete mixing distribution; these models have two mass points, with Gateaux derivatives used to test the whether additional mass points would provide a better fit. This alternative assumption makes little difference to the estimated association between SES and probability of transition.

I estimate regression models of the form:

$$\log(-\log(1-d_{it})) = \alpha(age) + \beta \mathbf{x}_{it} + \nu_i \tag{6}$$

where ν is an individual-level error term, which is assumed to be normally-distributed:

$$\nu \sim N(0, \sigma_\nu^2) \tag{7}$$

and uncorrelated with the explanatory variables:

$$Cov(\nu_i, \mathbf{x}_{it}) = Cov(\varepsilon_{it}, \mathbf{x}_{it}) = 0 \tag{8}$$

I estimate models including the same variables as in the main body of the paper (other than the addition of a random effect). I estimate these models using adaptive quadrature with 8 integration points, making use of the software GLLAMM (Rabe-Hesketh and Skrondal, 2006). This allows me to include individual-level random effects, while still with accounting for the complex survey design of the data (most notably the sampling and attrition weighting scheme, and the clustering of standard errors at the school-level).

C.1 Regression tables

The results of these models are reported in regression tables similar to those in Appendix A. Models for M0 are not reported, as these would not reliably converge. This would seem to be due to an over-reliance on the random effects to explain differences between individuals in this model with very few explanatory variables.

In addition to what is reported for models without random effects, the tables also show the estimated variance of the random effect and the results of a likelihood ratio test of the difference between the model and the counterpart model with no random effect. In each case, the model that accounts for unobserved heterogeneity does provide additional explanatory power.

The models for transition from 'likely to unlikely' are reported in Table 13, while the models for 'unlikely to likely' are reported in Table 14. This analysis provide broadly similar evidence on the association between SES and probability of transition as models in the main body of the thesis. However, there is a somewhat different pattern of association between age and probability of transition after accounting for unobserved heterogeneity between individuals.

Table 13: Estimated effects on risk of transition from reporting being 'likely' to apply to university to reporting being 'unlikely' to apply to university: hazard ratios

	M1	M2	M3	M4	M4C	M5	M5C
Age 16	1.24	1.21	1.24	1.23	1.24	1.23	1.23
	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Age 17	1.10	1.16	1.22	1.27	1.28	1.34	1.36
	(0.06)*	(0.07)**	(0.08)***	(0.08)***	(0.08)***	(0.09)***	(0.09)***
SES Q1 (Low)	1.77	1.81	1.20	1.17		1.19	
	(0.16)***	(0.17)***	(0.11)**	(0.10)*		(0.10)*	
SES Q2	1.66	1.46	1.25	1.23		1.22	
	(0.15)***	(0.12)***	(0.10)***	(0.10)***		(0.10)**	
SES Q4	0.67	0.73	0.77	0.77		0.77	
	(0.06)***	(0.06)***	(0.06)***	(0.06)***		(0.06)***	
SES Q5 (High)	0.22	0.30	0.41	0.41		0.43	
	(0.02)***	(0.03)***	(0.04)***	(0.04)***		(0.04)***	
SES Z-Score					0.65		0.66
					(0.03)***		(0.03)***
Male		1.75	1.76	1.71	1.72	1.71	1.72
		(0.10)***	(0.10)***	(0.10)***	(0.10)***	(0.10)***	(0.10)***
Ethnicity: Mixed		0.51	0.52	0.51	0.51	0.51	0.50
		(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Ethnicity: Indian		0.11	0.10	0.10	0.09	0.10	0.10
		(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Ethnicity: Pakistani		0.17	0.13	0.13	0.12	0.13	0.12
		(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Ethnicity: Bangladeshi		0.16	0.17	0.17	0.15	0.17	0.15
		(0.03)***	(0.04)***	(0.04)***	(0.03)***	(0.04)***	(0.03)***
Ethnicity: Black Caribbean		0.26	0.17	0.17	0.17	0.17	0.17
		(0.05)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
Ethnicity: Black African		0.12	0.10	0.10	0.09	0.10	0.09
		(0.03)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
Ethnicity: Other		0.18	0.16	0.16	0.15	0.16	0.15
		(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Attended Independent School		0.22	0.20	0.23	0.23	0.23	0.24
		(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.07)***
Attended Grammar School		0.15	0.33	0.34	0.34	0.35	0.35
		(0.03)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***	(0.07)***
Attended School with Sixth Form		0.81	0.83	0.84	0.84	0.83	0.84
		(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***	(0.05)***
Experienced workless household		0.96	0.87	0.83	0.74	0.84	0.76
		(0.08)	(0.07)	(0.07)**	(0.06)***	(0.07)**	(0.06)***
Ever experienced family separation		0.99	0.99	0.96	0.97	0.97	0.98
		(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Local Youth Unemployment Rate / 10		0.97	0.96	0.97	0.96	0.97	0.96
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
KS2 Z-Score			0.44	0.47	0.48	0.48	0.48
			(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
KS4 Z-Score (After results)				0.59	0.59	0.42	0.53
				(0.04)***	(0.04)***	(0.06)***	(0.04)***
KS4 Z-Score * SES Q1						1.72	
						(0.32)***	
KS4 Z-Score * SES Q2						1.65	
						(0.30)***	
KS4 Z-Score * SES Q4						1.22	
						(0.23)	
KS4 Z-Score * SES Q5						0.98	
						(0.24)	
KS4 Z-Score * SES Z-Score							0.78
							(0.06)***
Geographical		√	√	√	√	√	√
Number and order of siblings		√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√
χ^2 test of difference from previous model	.	667.42	397.97	63.32	183.09	13.76	12.53
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.01	0.00
Variance of Random Effect	2.19	1.64	1.33	1.31	1.34	1.29	1.31
LR test of diff. from non-RE model (χ^2)	385.48	271.36	231.84	241.99	253.91	232.91	241.32
p-value of above test statistic	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Number of individuals	9,247	9,247	9,247	9,247	9,247	9,247	9,247

Notes: Reporting hazard ratios. Standard errors (clustered by individual's school) in parentheses. Weighted using Wave 2 survey design and non-response weights. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.

Table 14: Estimated effects on risk of transition from reporting being 'unlikely' to apply to university to reporting being 'likely' to apply to university: hazard ratios

	M1	M2	M3	M4	M4C	M5	M5C
Age 16	1.04	1.06	1.04	1.03	1.03	1.03	1.03
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)
Age 17	0.74	0.73	0.73	0.89	0.89	0.88	0.87
	(0.05)***	(0.05)***	(0.05)***	(0.07)	(0.07)	(0.07)*	(0.07)*
SES Q1 (Low)	0.70	0.64	0.76	0.78		0.75	
	(0.07)***	(0.07)***	(0.08)***	(0.08)**		(0.08)***	
SES Q2	0.84	0.84	0.89	0.89		0.88	
	(0.07)**	(0.07)**	(0.07)	(0.07)		(0.07)	
SES Q4	1.37	1.34	1.20	1.18		1.17	
	(0.14)***	(0.13)***	(0.11)*	(0.11)*		(0.11)*	
SES Q5 (High)	2.54	2.42	2.01	1.95		1.96	
	(0.31)***	(0.28)***	(0.23)***	(0.22)***		(0.22)***	
SES Z-Score					1.36		1.37
					(0.06)***		(0.06)***
Male		0.52	0.53	0.54	0.54	0.54	0.54
		(0.03)***	(0.03)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***
Ethnicity: Mixed		1.80	1.81	1.80	1.78	1.82	1.80
		(0.32)***	(0.31)***	(0.30)***	(0.30)***	(0.30)***	(0.30)***
Ethnicity: Indian		4.19	4.85	4.76	4.85	4.77	4.81
		(0.93)***	(0.96)***	(0.96)***	(0.98)***	(0.96)***	(0.96)***
Ethnicity: Pakistani		5.55	6.85	6.56	6.93	6.58	6.84
		(0.97)***	(1.22)***	(1.15)***	(1.23)***	(1.15)***	(1.21)***
Ethnicity: Bangladeshi		7.76	8.25	7.75	8.14	7.85	8.16
		(1.46)***	(1.50)***	(1.40)***	(1.47)***	(1.41)***	(1.46)***
Ethnicity: Black Caribbean		3.87	4.51	4.35	4.21	4.42	4.24
		(0.82)***	(0.96)***	(0.90)***	(0.87)***	(0.91)***	(0.87)***
Ethnicity: Black African		7.92	10.20	9.62	9.92	9.75	9.84
		(2.37)***	(2.88)***	(2.69)***	(2.83)***	(2.75)***	(2.80)***
Ethnicity: Other		4.35	5.02	5.03	5.18	5.04	5.11
		(0.96)***	(1.21)***	(1.17)***	(1.17)***	(1.18)***	(1.17)***
Attended Independent School		1.38	1.54	1.48	1.51	1.46	1.49
		(0.61)	(0.63)	(0.57)	(0.59)	(0.56)	(0.58)
Attended Grammar School		2.06	1.08	1.03	0.98	1.00	0.97
		(0.61)**	(0.32)	(0.30)	(0.29)	(0.29)	(0.29)
Attended School with Sixth Form		1.09	1.06	1.05	1.05	1.05	1.05
		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Experienced workless household		0.98	1.04	1.09	1.12	1.08	1.11
		(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Ever experienced family separation		1.09	1.09	1.09	1.09	1.10	1.10
		(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Local Youth Unemployment Rate / 10		1.05	1.03	1.03	1.03	1.03	1.03
		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
KS2 Z-Score			1.71	1.59	1.58	1.58	1.58
			(0.07)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***
KS4 Z-Score (After results)				1.78	1.78	1.97	1.96
				(0.13)***	(0.13)***	(0.35)***	(0.17)***
KS4 Z-Score * SES Q1						0.79	
						(0.16)	
KS4 Z-Score * SES Q2						0.90	
						(0.21)	
KS4 Z-Score * SES Q4						1.53	
						(0.46)	
KS4 Z-Score * SES Q5						0.63	
						(0.17)*	
KS4 Z-Score * SES Z-Score							1.20
							(0.10)**
Geographical		√	√	√	√	√	√
Number and order of siblings		√	√	√	√	√	√
Months of birth and interview	√	√	√	√	√	√	√
χ^2 test of difference from previous model	.	334.66	210.08	61.99	123.07	9.78	4.56
p-value of above test statistic	.	0.00	0.00	0.00	0.00	0.04	0.03
Variance of Random Effect	1.51	1.28	1.04	0.99	0.99	1.00	0.99
LR test of diff. from non-RE model (χ^2)	178.19	144.84	111.63	101.51	100.05	101.96	100.23
p-value of above test statistic	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Number of individuals	5,330	5,330	5,330	5,330	5,330	5,330	5,330

Notes: Reporting hazard ratios. Standard errors (clustered by individual's school) in parentheses. Weighted using Wave 2 survey design and non-response weights. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimated risks are relative to the following base categories: Age 15, SES quintile group 3, attended a non-selective state school, white, and female. Tests of model fit are relative to the model one column to the left, with the following exceptions: M4C is relative to M3, M5 is relative to M4, and M5C is relative to M4C.