Department of Quantitative Social Science

What's the link between household income and going to university?

Jake Anders

DoQSS Working Paper No. 12-01
March 2012

Leading education and social research
Institute of Education
University of London

DISCLAIMER

Any opinions expressed here are those of the author(s) and not those of the Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

DoQSS Workings Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Department of Quantitative Social Science. Institute of Education, University of London. 20 Bedford way, London WC1H 0AL, UK.

What's the link between household income and going to university?

Jake Anders[*†]

Abstract. The association between household income and university entry is a matter of clear academic and policy interest. This paper sheds new light on the matter using the LSYPE, a recent longitudinal survey from England. While those in the top income quintile group are more likely than those in the bottom quintile group to attend university (66% vs. 24%), much of this gap is explained by earlier educational outcomes. The paper also examines admissions decisions in more detail, separating applying from attending. This analysis yields results suggesting most of the difference in participation rates is driven by the application decision. The attendance gap conditional on having applied is much smaller (85% vs. 68%) and closes completely when earlier educational outcomes are taken into account. Finally, the paper considers attendance at high quality Russell Group universities. By contrast with the main analysis, the Russell Group attendance gap persists even among those who attend university. The findings suggest policies aimed at reducing the university participation gap at point of entry face small rewards. More likely successful are policies aimed at closing the application gap, for example encouraging a wider cross-section of the population to apply and ensuring they have the necessary qualifications.

JEL classification: Higher Education, Household Income, Socioeconomic Gradient, Intergenerational Mobility.

Keywords: I24, J62.

[*]Institute of Education, University of London. E-mail: jake@jakeanders.co.uk.

[†]This research forms part of the Nuffield Foundation project "Higher Education Funding and Access: exploring common beliefs". The project is based at the Institute of Education and the Institute for Fiscal Studies. It is directed by Claire Crawford, Lorraine Dearden, Alissa Goodman, John Micklewright and Anna Vignoles. Thanks to my supervisors Lorraine Dearden and John Micklewright for all their help. Thanks also to Claire Crawford and John Jerrim for their feedback and to Peter Backus for his advice on local polynomial smoothing.

1 Introduction

The association between household income and young people's university entry is a matter of clear academic and policy interest. Despite previous studies into the matter suggesting that much of this gap emerges significantly earlier in a young person's life, we see regular policy ideas aimed at reducing the size of the gap through changes to the university admissions process. One example is the suggestion that universities use more contextual data on applicants contained in the government's Social Mobility Strategy (HM Government, 2011). In this paper I make three important contributions to this matter.

First, I demonstrate the extent of differences in university participation by household income for a very recent cohort, using a large longitudinal dataset with reasonable income measurement, rich socioeconomic background information, school characteristics and attainment data. This is in stark contrast to much previous research, where available income data have been of low quality or not present at all. Many studies have had to make do with poor categorical substitutes, such as parents' housing tenure (Gayle et al., 2003). This has prevented analysis which can look across the income spectrum.

Second, I examine the relationship between household income and the decision to apply to university, as well as the actual act of attending. Hence, the likelihood of attending conditional on having applied, as household income varies, can be isolated. I use the more in depth information on the university application process available in the dataset in order to achieve this. It delivers a novel and interesting addition to the body of knowledge.

Third, I analyse the differences in participation rates by household income for differing quality universities, exploiting the large size of the dataset, and the fact that the young people interviewed were asked which university they attend. Along with the general concern at the participation gap by income, more specific concerns have been raised that although participation in general has increased, those from worse-off socioeconomic backgrounds are disproportionately attending lower prestige institutions. Hence their returns from a university education may be lower (Chevalier and Conlon, 2003).

Much previous research uses datasets which refer to cohorts passing through the education system at least two decades ago. The education system in general, and higher education (HE) in particular, has changed dramatically since that time. Greatly increased overall HE participation rates, particularly beginning in the early 1990s, and the introduction of tuition fees in 1998 both diminish the relevance of findings from these data for today's world. In this paper, I use the Longitudinal Study of Young People in England (LSYPE) (Department for Education and National Centre for Social Research, 2011), whose participants reached age 18 in the academic year 2008-09. While the dataset does have its drawbacks, which I shall discuss later, it also has many strengths.

In Section 2, I begin by briefly summarising the economic theory and previous research which suggests a link between household income and HE participation going beyond correlation. I also consider previous UK evidence on socioeconomic status and university attendance. From this theoretical and empirical grounding, in Section 3, I outline some of the features of the LSYPE particularly pertinent to my analysis. I explain the rationale of modelling university admissions as sequential decisions in Section 4. I then examine the patterns we see between the household income and participation decisions in Section 5. This leads on to analysis using standard regression techniques in Section 6, drawing lessons from the literature on the specification of models of the determinants of children's attainments, to control for potentially confounding factors and evaluate the remaining household income gradients.

In Section 7, I turn my attention to differences in participation rates across universities of differing prestige and, by proxy, quality. The probability of attending a Russell Group university, as distinct from other universities, is assessed in order to see if household income seems to be associated with differences in the quality of institution attended.

2 Theory and previous research

This paper is influenced by more than one body of economic literature. It draws on work to isolate the causal impact of household income on a child's economic outcomes (Blau, 1999) and the empirical literature aiming to identify the relationship between socioeconomic characteristics and university attendance (Haveman and Wolfe, 1995; Blanden and Gregg, 2004).

Human capital theory as set out in Becker and Tomes (1986) develops a model of the transmission of earnings from parents to their children under the assumption that parents maximise utility, but care for their children. Under further plausible assumptions of either capital constraints or the inclusion of their children's human capital in a parent's utility function the model predicts a direct effect of parental earnings on children's earnings (Becker and Tomes, 1986, p. 12). So long as at least some of this earnings effect is transmitted through increased education for the children of those with higher incomes, it seems likely that we will encounter the effect in university participation models. The signalling hypothesis sees education in part as a signal to future employers of higher levels of innate, unobserved ability (Spence, 1973). However, educational signalling is by definition costly, hence there is potential for individuals to be credit constrained when choosing how much to purchase under this theoretical perspective too. If educational decisions are credit constrained, higher levels of household income ease the constraints on either investing in human capital or purchasing costly education as a signal to employers to achieve higher future earnings. As such, once the credit constraint no longer binds we would expect no additional marginal benefit from household income.

In using the term credit constraints I refer primarily to its long run concept, as distinguished by Carneiro and Heckman (2002, p.705-706). While there may be short run credit constraints at play [1], use of a measure of permanent income means that this paper remains silent on this matter. Unfortunately, in the absence of an exogenous household income shock, it is impossible to separate out the causal impact of permanent income from the unobserved characteristics it is correlated with. Credit constraints on investing in human capital exist because one cannot realistically use one's future productivity as collateral: there exist both ethical issues, given its equivalence with selling oneself into slavery, and practical issues, given the asymmetric information problem with respect to one's future employment plans. Richer parents will ease these credit constraints by giving (or perhaps cheaply lending) to their offspring. They are also more likely to encourage them to stay in education, thereby not imposing the indirect costs that a lack of support leads to (Atkinson, 1983, pp. 116-118).

Alternative theoretical backing for broadly the same effects may be drawn from the more sociological and psychological 'good parent theory' (Mayer, 2002, p. 13). This provides a more indirect approach to the potential association between income and educational suc-

[1] Dearden et al. (2004) suggest that there is only evidence that credit constraints exist for a small part of the population.

cess, suggesting that a lack of income negatively impacts on parenting quality in some way and that it is this that closes off educational opportunities to the child. Variants include the 'parental stress' version, suggesting that the stress caused by low income impairs the ability to be a good parent, and the 'role model' theory, suggesting that lower income parents either have or develop "values, norms, and behaviours that are 'dysfunctional' for success in the dominant culture" (Mayer, 2002, p. 15).

Sociologists also argue for class-based explanations of educational differentials. An influential theory for this is known as the Breen-Goldthorpe model of relative risk aversion (Breen and Goldthorpe, 1997). It postulates 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 suggests that individuals from different classes will have, on average, different educational aims and this will lead to different behaviour in choice under uncertainty. Given the focus on household income in this paper, I do not include a measure of social class, such as occupational status, in my analyses. In doing so I am not arguing that explanations such as these are unimportant. However, given the likely correlation between social class and income, coefficients for household income will pick up these sociological effects.

This paper does not estimate the specific causal impact of household income on university attendance. However, it is nevertheless informative to consider the attempts that have been made to do so. Much of this literature comes from the US, where institutional differences mean that evidence does not directly provide guidance on the UK. Blau's (1999, p.263) summary, echoed in much the same form by Jenkins and Schluter (2002), of much of the US literature on the role of household income with regard to educational outcome includes five key points: permanent income is more important than transitory income; conditional income effects are smaller than unconditional income effects; income effects are small compared to other parental characteristics; income effects differ by stage of childhood; and income effects are non-linear. All of these have some bearing on this paper, I consider them in turn.

- *Permanent and transitory income* - It is perhaps of little surprise that permanent income is seen to have a larger effect on children's outcomes, compared to transitory income.

As Susan Mayer points out "families can smooth consumption over short periods of low income by either borrowing, using savings, or calling on family, friends, charity, or public services to smooth their living standards" (Mayer, 2002, p.21). More methodologically, transitory income has a larger variance than permanent income and correlations between it and outcomes of interest will be smaller for this reason. Nevertheless, transitory income at the point of key decisions in a child's development can have a separate impact if it results in the decisions being credit constrained. I consider this matter in another paper (Anders et al., 2012). In this paper I follow Blau's definition of permanent income (Blau, 1999, p.263) by averaging across the LSYPE's income measures in waves 1 to 4.

- *Conditional income effects* - Conditional income effects, by which is intended the remaining impact of income once other socioeconomic variables have been included in the model, are likely to be smaller than unconditional income effects because of the correlation between household income and other observable socioeconomic factors such as parental education. In attempting to uncover the causal impact of household income, Blau notes the danger of including variables that may be jointly chosen with household income, including in this factors such as household structure and parental education levels. I do include such correlates in some models, because although it could downward bias estimates of income effects, comparison of the models suggests potential causal routes through which income effects may be working.

- *Size of income gradients* - I will consider the relative size of income gradients and other family characteristics. Previous research has generally suggested that other factors, such as parental education, are relatively more important in themselves.

- *Income gradients by age* - The scope of this paper does not extend to considering the effects of income across different stages of childhood. Any associations identified only relate to the years under consideration. For most models this means that any income gradients only inform us about the teenage years, since prior attainment captures socioeconomic variation up to that point.

- *Non-linear income effects* - Before undertaking parametric modelling, I investigate the potential for non-linear associations between income and university admissions. It makes intuitive sense that this may well be the case since, as Mayer puts it, "[m]ost

> people think that an extra $1,000 helps a family with $10,000 a year more than it helps a family with $100,000 a year." (Mayer, 2002, p.25)

Blau (1999) goes on to exploit fixed effects at the grandparent, parent and child levels to attempt to control for the unobserved heterogeneity that results in income endogeneity and hence upward biased estimates of the causal impact of income. He still identifies effects on child development, albeit ones that "are too small to make income transfers a feasible approach to achieving substantial improvements in development outcomes of low-income children" (Blau, 1999, p.273), also pointing out that his evidence seems to confirm that other child and family characteristics seem to have larger effects. Shea (2000) attempts to isolate the causal impact of parental income on their child's future earnings by considering changes in parental education which are down to 'luck' - proxied by father's union, industry and job loss due to company closure. His results suggest no significant relationship between changes in income of this type and child's future earnings.

One paper that does directly address the question of college enrolment is Acemoglu and Pischke (2001). They argue that changes in the overall income distribution in the US generate the exogenous shifts in household income required to identify its causal impact. Hence, by examining the shifts in college enrolment across the income distribution during the same period they argue that the causal impact of household income may be identified. These estimates imply that a "10 percent increase in family income [will] increase college enrolments by 1-1.4 percentage points" (Acemoglu and Pischke, 2001, p.903). They also compare these causal estimates with estimates that will include wider family background effects suggesting that "family income, rather than other factors related to family background, explain[s] 27 percentage points of the 36 percentage point difference in the enrolment rates of children from the bottom and top quartiles in 1992" (Acemoglu and Pischke, 2001, p.901). Finally, by allowing the impact of income to vary across income quartiles they present suggestive results that "indicate that even relatively rich families may not be completely unconstrained. In addition, income may matter for reasons other than credit market constraints, for example, because college is, to some degree, a consumption good rather than a pure investment good" (Acemoglu and Pischke, 2001, p.903). It is, however, pointed out that statistical significance is much weaker, because of the effective quartering of sample size for each quartile group estimate that this entails.

Previous UK empirical studies have also suggested household income's link with higher levels of university attendance. This relationship has proved robust to other controls in some studies. Blanden and Machin (2004) use several cohorts of data spanning very different HE cohorts to examine the changing relation between household income and university attendance as the proportion of the population who attend university greatly increased. They use several methods, with differing measures of participation and income inequality, to show that the expansion of participation has not been equally distributed across the population. Rather it has disproportionately resulted in increased participation rates among young people from better off families.

Gayle et al. (2003) use a single cohort of the Youth Cohort Study (YCS) to model demand for HE. After controlling for prior attainment their models suggest that ethnicity, housing tenure, region and parental education continue to show a continued association with the probability of HE attendance. They argue that in the absence of income measurement from the YCS, we can view housing tenure as a proxy for parental wealth, and hence that their results are suggestive of a continued role for this. On the contrary, Marcenaro-Gutierrez et al. (2007) also use the YCS, but this time take advantage of multiple cohorts between 1994 and 2000 to analyse the socioeconomic gradients associated with the probability of attending university. In their analysis they identify no association between socioeconomic factors and the probability of attending university once they condition on academic attainment at 16 or 18. They hence conclude that the socioeconomic inequality in university attendance arises earlier in the education system.

The issue of high levels of attrition from the YCS causes problems for the inference from its estimates to the population. This has in turn led many researchers to seek alternative data sources. Chowdry et al. (2010) use administrative data, formed by linkage of the National Pupil Database (NPD) and Higher Education Statistics Authority (HESA) data, to consider the association between an index of socioeconomic status and HE attendance. Before their main analysis, they demonstrate the policy importance of gaps in HE attendance by considering the gap by whether an individual is eligible for free school meals (FSM). They show that, for their cohorts, "only 14% of [state school] pupils who are eligible for free school meals participate in higher education at age 19/20, compared with 33% of pupils who are not eligible for free school meals, a very large gap indeed" (Chowdry et al., 2010, p.2). Although the figures are not comparable, this same gap is noticeable in the LSYPE, albeit with higher overall university attendance. In the LSYPE, 39% of non-FSM state school pupils attend university, compared to 20% of those who are FSM eligible. I am aware of the imperfection of using FSM eligibility in this context (Hobbs and Vignoles, 2010) and use it only for comparison purposes.

In Chowdry et al.'s main analysis they demonstrate a raw gap in the probability of university attendance between top and bottom socioeconomic quintiles of 40.7 percentage points for boys and 44.6 percentage point for girls. They also use linear probability regression models with school fixed effects to estimate the remaining socioeconomic gap controlling for other factors. The gap between the top and bottom quintiles is significantly reduced once other individual and school controls are included, with the gap standing at 29.9 percentage points for boys and 35.8 for girls. This is reduced still further once prior attainment is controlled for, first at age 11 with the gap at 21.1 percentage points for boys and 25.6 for girls, then at age 16 with gaps of 8.7 percentage points for boys and 11.3 percentage points for girls. (Chowdry et al., 2010, p.15)

A major determinant of application and attendance at university will be an individual's underlying ability. Unfortunately, we cannot directly control for this. If ability were not correlated with socioeconomic background (such as household income) this should not bias estimates. However, because a non-zero correlation seems likely (Haveman and Wolfe, 1995, p.1833) it is necessary to include a proxy for ability, such as prior attainment, to reduce omitted variable bias. However, this does have drawbacks. As household income likely does impact upon intermediate attainment outcomes, only the additional impacts of income (after the prior attainment measure but before applying to university) will then be identified by a model including this attainment measure.

3 Data

The LSYPE is a major longitudinal study funded by the Department for Education. It is made up of seven 'waves' conducted annually, beginning in Summer 2004 when cohort members were in Year 9 (aged 13-14). Interviews were conducted with young people and their parents, covering information about the cohort members themselves and their households.

This is then linked with administrative data, in the form of the National Pupil Database (NPD), to provide results for statutory examinations beginning with Key Stage 2 SATS, taken at the age of 10 or 11 in the last year of primary school.

More in depth analysis of the data issues raised here, and use of the LSYPE for modelling HE participation more generally, may be found in a companion paper (Anders, 2012).

Wave 7, which is the final wave, covers young people aged 19-20. Hence the data allow us to model the entry to university through what might be thought of as the 'traditional' route, going from sixth form or further education college to university, either the same year or after a single gap year. While this includes the majority of those who attend university, the exclusion of a potentially interesting subpopulation should be noted. Furthermore, previous analysis has suggested that those from poorer backgrounds are more likely to enter HE later, so this research could overstate the income gap (Bekhradnia, 2003, p.2).

Table 1: Percentages of Young People Achieving Key Application Milestones

	Overall	**Female**	**Male**
University Applicant	50.5 (0.89)	54.2 (1.16)	46.5 (1.11)
Sample size	*8677*	*4456*	*4187*
University Attendee	38.6 (0.84)	41.9 (1.15)	35.0 (1.01)
Sample size	*8677*	*4456*	*4187*
HE Attendee	43.3 (0.87)	46.6 (1.15)	39.8 (1.06)
Sample size	*8677*	*4456*	*4187*
Russell Group Attendee	9.5 (0.48)	10.2 (0.70)	8.6 (0.63)
Sample size	*8665*	*4450*	*4181*
Conditional University Attendee	76.5 (0.75)	77.4 (1.00)	75.3 (1.11)
Sample size	*5323*	*2889*	*2410*
Conditional Russell Group Attendee	24.6 (0.98)	24.4 (1.38)	24.4 (1.42)
Sample size	*4191*	*2277*	*1870*

Notes: Standard errors in parentheses. Weighted using Wave 7 LSYPE Weights, which attempt to adjust for oversampling and attrition. Application, Offers, Acceptances and Attendance calculated across Wave 5, 6 and 7. Sample: All Wave 7 respondents.

Table 1 shows the percentage of individuals who reach the milestones in the university application process that we will be analysing, specifically applying, attending and attending a Russell Group university. Also included, for comparison purposes, is the proportion who undertake HE. This is a broader definition than university, including those taking HE courses at Further Education colleges. One weakness of the LSYPE appears to be that these proportions are rather higher than we would anticipate from comparison with administrative data. For example, the Higher Education Initial Participation Rate (HEIPR) for ages

17-19 in 2008/09 is 32.9% and for 2009/10 is 34.1% (Department for Business, Innovation & Skills, 2011). Since our LSYPE admissions measurement in fact spans these two years we would expect our estimate of HE attendance to fall somewhere between these two figures, whereas in fact it is notably larger at 43.3%.

I assume that this is related to attrition from the study. While attrition in the LSYPE is reasonable by the standards of similar studies (62.4% of the initial sample remain by Wave 7) it still has the potential to bias sample estimates. Non-response is modelled prior to release by either the Department for Education or the National Centre for Social Research and used to produce weights to reduce this bias. However, it seems likely that the probability of attrition is correlated with unobservable factors and hence not possible to model completely accurately.

The LSYPE measures household income at each wave between 1 and 4, however the methods vary across the waves. Since our interest is in the underlying permanent income (Jenkins and Schluter, 2002, p.2), an approximation to this is calculated by averaging across the waves. By allowing the permanent income measure to include those with measurement from fewer than all four waves this also helps reduce our missing data problem, albeit at the cost of perhaps being further from true permanent income in those cases. Summary statistics of the household equivalised income are shown in Table 2. Income has been equivalised by dividing by the square root of household size.

Table 2: Household Equivalised Income Summary Statistics

Characteristic	**Value**
Mean	18,511
Standard Deviation	13,508
Minimum	261
Maximum	174,010
1st Percentile	2,959
10th Percentile	5,904
Median	15,097
90th Percentile	36,534
99th Percentile	60,110
N	7,344

Notes: Incomes adjusted to Wave 1 (2004) prices using Annual RPI. Equivalised by dividing by square root of family size. Weighted using LSYPE Wave 7 respondent weights. Sample: Wave 7 respondents with valid income data from at least one of Waves 1-4.

Overall, the LSYPE appears to underestimate household incomes relative to fairly comparable figures in other surveys. Comparison with the Family Resources Survey suggests underestimation, particularly below £60,000 gross household income. This may be due to failure to take into account certain kinds of benefit income, notably in work tax credits which cut out at this point (Anders, 2012, p.26).

The LSYPE comes linked to selected elements of the National Pupil Database (NPD). This provides information on the young people's schooling experience and attainment data from Key Stage 2 (age 11), Key Stage 3 (age 14) and Key Stage 4 (age 16). Having high quality data on prior attainment with low non-response is a major advantage compared to many previous survey studies. At present Key Stage 5 results are not included in the standard LSYPE release, however I would not want to use these as part of most of my models in any case. Firstly, unlike Key Stage 4 qualifications, A Levels etc. are taken after the end of compulsory schooling, so would only be available for a self-selected sample. Furthermore, A Level results are deeply embedded in the university application process, particular with pre-results application.

Some of the SATS attainment data from Key Stages 2 and 3 are missing where an individual was not in the state education sector and hence either did not take SATS or, if they did, the school chose not to report them. Independent schools are under no obligation to do either, although some 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[2].

For Key Stage 2 (KS2), I use the average raw point score across all three subjects (Maths, English and Science) in KS2 SATS. KS2 SATS are relatively low stakes examinations, although there is some limited use by secondary schools for streaming. 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.

For Key Stage 4 (KS4), I use the official capped GCSE score. GCSEs (General Certificate of Secondary Education) are high stakes public examinations, taken at the end of compul-

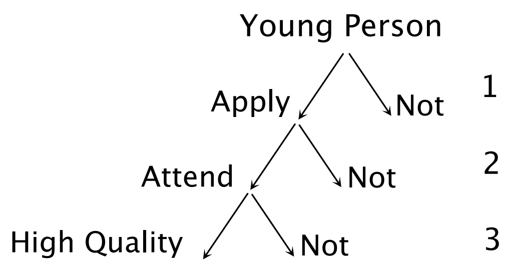
[2] It it also worth noting that Key Stage 3 SATS were abolished in England in 2008.

sory 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 equivalents. This is in contrast to the uncapped score, which uses all GCSEs and equivalents taken and hence is more subject to manipulation by schools.

4 University admissions as sequential decisions

Previous research has considered differences in university participation by their socioeconomic characteristics. However, the story is more complicated: the process of university admissions is a set of sequential decisions. Although there are in fact many nuances to this model, and many more hurdles in the process, I have chosen to simplify these into three steps: application, attendance, and attendance at a high-status university (see Figure 1).

Figure 1: Simplified Admissions Model



My decision to simplify in this way was made for reasons both of clarity and the limitations of the data. In exploring the data I discovered that very few applicants fail to receive any offers and very few of those offered a place do not accept any of them. The questions in the LSYPE then do not allow us to distinguish between those who do not attend due to failing to fulfil their conditional offers and those who choose not to attend for some other reason.

Nevertheless, assumption of even a simple sequential model like this allows benefits in interpretation. In particular, in addition to models of the probability of applying to university and the probability of attending university, we can also consider the probability of attending university conditional on having applied. Given that one must apply in order to attend, the probability of attending conditional on having applied is related to the probabilities of applying and attending by the law of conditional probability, as shown in Equation 1.

$$P(\text{Attend}|\text{Apply}) = \frac{P(\text{Attend} \cap \text{Apply})}{P(\text{Apply})} \tag{1}$$

While this seems to be a relatively unusual approach, some precedent is provided by Gayle et al. (2000), who use non-nested models to show that girls are more likely than boys to continue in education past the age of 16, but conditional on having done so boys are more likely than girls to continue into HE.

A major consideration in its use is, however, the problem of selection. Those who apply to university are, of course, not a random sample of the population. They self-select presumably because it is expected utility maximising to do so. In this case I take this to mean two things. Firstly, they believe that, for them, university education has net present benefits greater than costs. Secondly, that they have some chance of receiving an offer and fulfilling any conditions required to take that offer up. Applying to university is not costless, either in terms of money or time. We cannot, therefore, consider parameter estimates from conditional attendance models to apply to the whole population, only to the subpopulation of individuals who have applied to university.

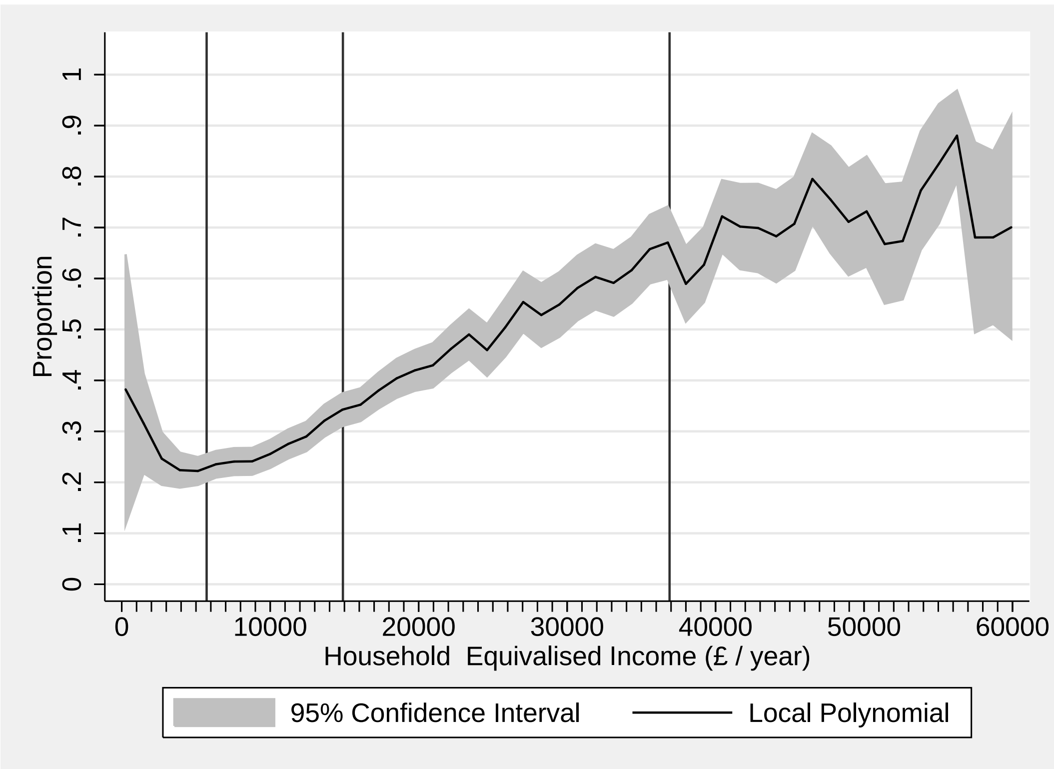
A second consideration is that the inevitably smaller sample size of the conditional models means that standard errors of estimates will be larger simply for this reason. This means that comparisons between the conditional and unconditional models on the basis of changes in significance are not reliable (Gayle et al., 2000, p.63). Significance testing of estimates in a model is still valid, but any comparisons between conditional and unconditional models must be on the basis of the substantive estimates, not their significance.

5 Analysis of the decision process

Looking at the raw relationship between household income and university admission (application, attendance and the conditional relationship) tells an interesting story in itself. In order to do this I employ the non-parametric technique of local polynomial smoothing. It allows me to assess the relationship without making any functional form assumptions that could distort the interpretation. I have chosen to estimate the appropriate bandwidth using the method suggested by Silverman (1986, p.48) to fit the local polynomial.

I first consider the simple unconditional university attendance model, comparable to much previous research in this area. Figure 2 demonstrates graphically the well known socioeconomic gap. The vertical lines on the graph show the 10th, 50th and 90th percentiles of equivalised income. University participation increases along with equivalised household income. For a large portion of the income distribution the relationship appears linear, however two features of the relationship seem particularly worthy of note.

Figure 2: Non-Parametric plot of Household Equivalised Income and University Attendance



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 1590.738. Sample size: 7791. Vertical lines show 10th, 50th and 90th percentiles of income.

First, at the very bottom of the distribution (below approximately £6,000) the relationship seems to begin in reverse: as household income rises participation rates initially fall. The average level of other typical socioeconomic status indicators (such as parental education level) are also lower in this bottom section of the income distribution, suggesting it is not purely driven by an unexpected profile of individuals in this very low income bracket. However, further investigation suggests it is related to differences in university attendance rates across different ethnic groups and measurement error of certain kinds of income in lone parent families.

Regardless, testing of the hypothesis of a different slope for the bottom section using a simple regression model specified with a discontinuity at this point suggests not allowing for a kink here: a formal Wald test of the difference between linear slopes across the two sections fails to reject the null hypothesis of no difference at the 5% level. There are few young people with household income in this bottom section, as witnessed by the large confidence intervals.

Second, the increase in attendance rate seems to plateau at about 75%. This corresponds with an equivalised income of roughly £40,000, somewhere around the 92nd percentile of the income distribution. This would appear to fit in with a story of credit constraints driving the relationship, at least in part. It also accords with previous evidence on the non-linearity of the relationship between income and children's outcomes (Mayer, 2002, pp.25-27)

Table 3: Probability of university application or attendance by income quintile group

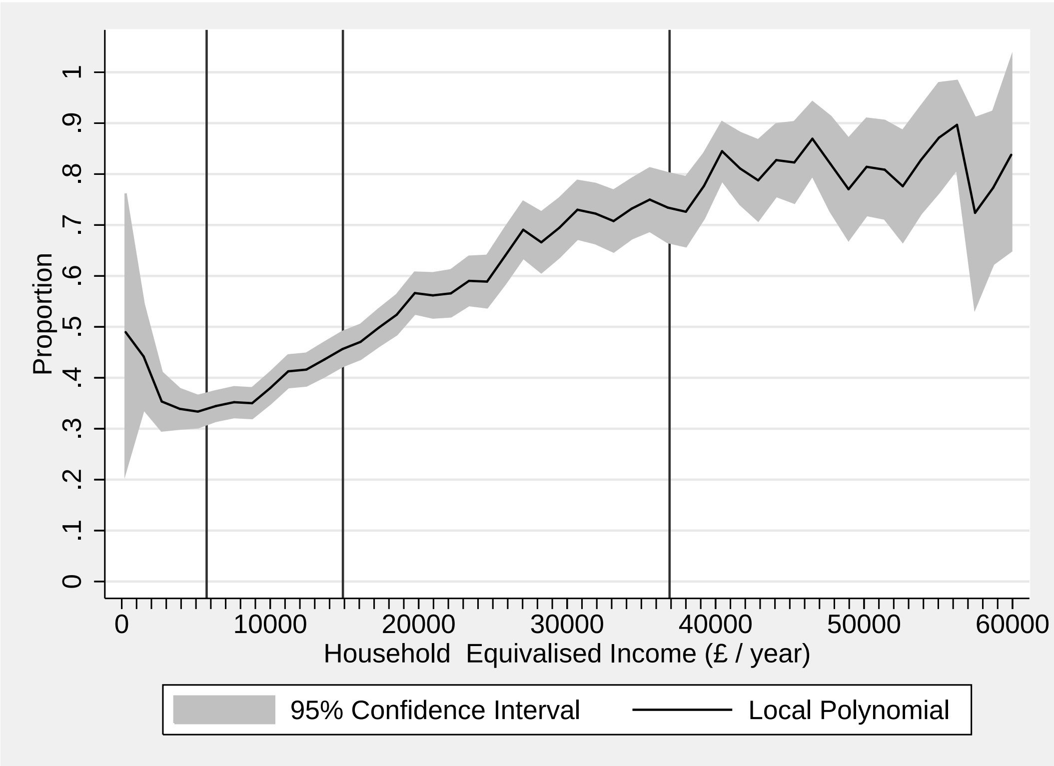
Variable	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**N**
University Application	0.35	0.38	0.47	0.57	0.77	7860
University Attendance	0.24	0.26	0.35	0.45	0.66	7860
University Attendance (Condition on Application)	0.68	0.68	0.74	0.79	0.85	4847

Notes: Adjusted using LSYPE Wave 7 respondent weights, which attempt to adjust for oversampling and attrition. Sample: Wave 7 Participants with valid responses for variables used in models.

For a simple indication of the participation gap (attendance) by household income I compare the attendance rate for those in the bottom income quintile with those in the top quintile. We see that those in the top income quintile group are 2.7 times as likely to attend university as those in the bottom quintile group ($\frac{0.66}{0.24}$). Results for all quintiles are shown in Table 3. While large, this is nowhere near as large as some estimates of the participation gap. An example of these is cited by Prime Minister David Cameron in a speech on HE reform, where he referred to a “status quo in which a person who is well-off is seven times more likely to go to university than someone from a poor background. This is the appalling situation we’re in” (Cameron, 2010).

In any case, this analysis alone tells us nothing about the point at which the gap emerges. One could, for example, take from this that young people from across the income spectrum are applying to university, but those with lower household incomes do not get places. However, as we are about to see, this interpretation is questionable.

Figure 3: Non-parametric plot of Household Equivalised Income and University Application

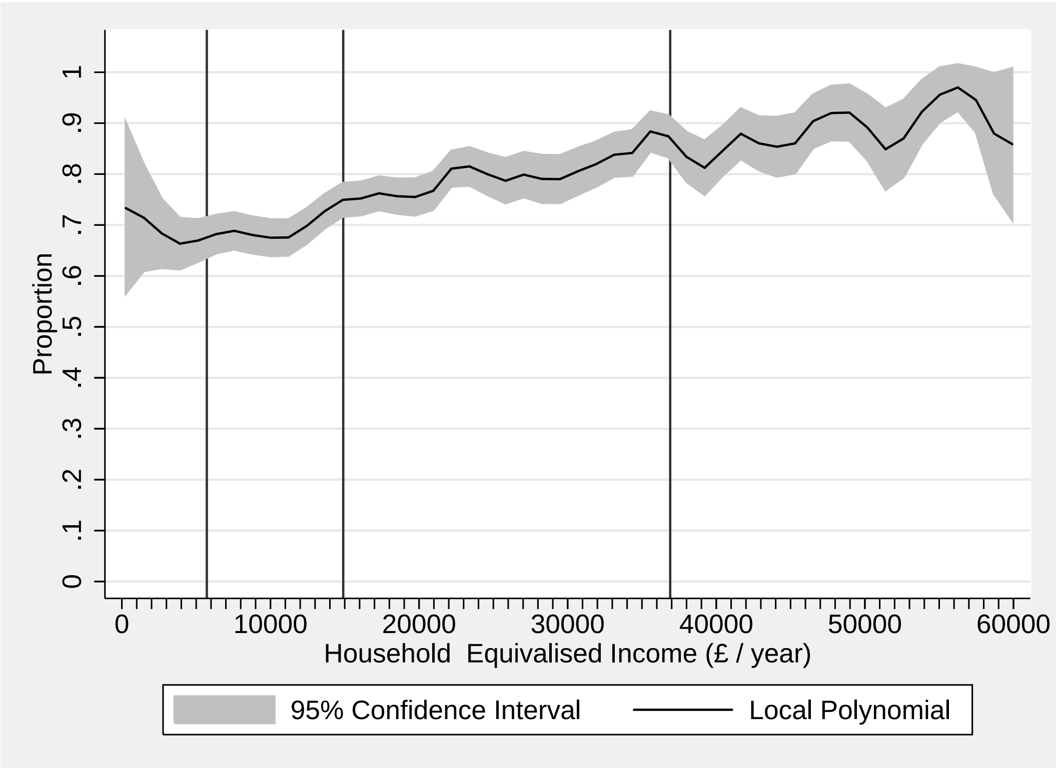


Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman’s optimal bandwidth of 1590.738. Sample size: 7791. Vertical lines show 10th, 50th and 90th percentiles of income.

Using the additional information available in the LSYPE, still looking at simple correlation, we turn our attention to the relationship between household income and university application, and by extension university attendance conditional on having applied. As Figure 3 shows, a strikingly similar relationship holds as for the unconditional attendance model. As such, a drastically reduced relationship between household income and conditional university attendance is seen throughout the income distribution, in Figure 4.

Instead of during the admissions process itself, the bulk of the raw gap appears to arise at or

Figure 4: Non-parametric plot of Household Equivalised Income and University Attendance, conditional on Application



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 2295.6094. Sample size: 4780. Vertical lines show 10th, 50th and 90th percentiles of income.

before the decision to apply. Once a young person has applied to university the probability of a person attending in the top quintile group is just 1.2 times larger than a person in the bottom quintile. This is before any confounding factors have been considered.

The extent to which this is self-selection on the basis of other characteristics cannot be identified by looking simply at this correlation. In order to begin understanding these, and the potential causes of the remaining gap for those who have applied, I turn to regression modelling.

6 Regression models of university admissions

I estimate regression models of university application (Apply), university attendance (Attend) and university attendance conditional on having applied (Conditional Attend). Given the binary nature of each of these decisions, I choose to use probit regression. This is preferable to using a linear probability model where, among other issues, there is no constraint on the predicted probabilities falling between 0 and 1 (Thomas, 2005, pp.445-450).

Separate models are estimated for males and females on the basis of the differences in relevant characteristics between the genders in the data, see Table 12, and in the previous literature, see Broecke and Hamed (2008). Significant differences in characteristics include university application and attendance rates, and attendance rates at grammar and independent schools. However, likelihood ratio tests of models excluding and including the possibility of different effects by gender fail to reject the null hypothesis of no difference. I therefore interpret differences between the separate models by gender with caution.

Several models also use what might be referred to as a 'value-added' specification, using a measure of the individual's prior attainment in the model. Estimates then aim to refer to the additional effect after the point prior attainment is measured. While it is clear that there are many drawbacks to this class of specification (Todd and Wolpin, 2004, p.7-9) the available data do not provide for more demanding specifications, such as the so called 'cumulative' specification. This specification would allow for, among other factors, the possibility of correlation between attainment measures and future family inputs.

$$
\begin{aligned}
A_i^* &= \alpha_0 + \beta_1 \text{Income≤£40,000} + \beta_2 \text{Income>£40,000} + \boldsymbol{\gamma}\mathbf{X} + \varepsilon_i \qquad (2)\\
A_i &= 1[A_i^* > 0]\\
&= 0[A_i^* \leq 0]
\end{aligned}
$$

$$
\begin{aligned}
\text{Hence } Pr(A = 1|\mathbf{X}) &= Pr(\varepsilon > -[\alpha_0 + \beta_1 \text{Income≤£40,000} + \beta_2 \text{Income>£40,000} + \boldsymbol{\gamma}\mathbf{X}])\\
&\text{where } \varepsilon \sim N(0,1) \text{ equivalently}\\
Pr(A = 1|\mathbf{X}) &= \Phi(\alpha_0 + \beta_1 \text{Income≤£40,000} + \beta_2 \text{Income>£40,000} + \boldsymbol{\gamma}\mathbf{X}) \qquad (3)
\end{aligned}
$$

The regression models estimated are generally of the form shown in Equation 2. A_i^* represents an unobserved index of propensity to apply to or attend university, while A_i represents the observed decision of whether or not to attend. α_0 represents the value of this index for a base set of characteristics; **X** represents a vector of additional controls varying by model and discussed below; and ε_i is the error term. I use a probit model, meaning I assume that this error term is normally distributed. Hence, the probability of a positive outcome is obtained by evaluating the outcome of the regression equation through Φ (the cumulative distribution function of the standard normal distribution), as shown in Equation 3.

Table 4: Regression Models

Variables	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**
Household Income	√	√	√	√	√	√	√
KS2 Attainment		√	√	√	√	√	√
KS4 Attainment				√	√		
Family Type			√		√	√	√
Parental Education			√		√	√	√
Region			√		√	√	
Ethnicity			√		√	√	√
Month of Birth			√		√	√	√
Sibling Effects			√		√	√	√
KS3 School Characteristics						√	
KS3 School Fixed Effect							√

The first model (M1) simply includes equivalised household income, modelled using a piecewise linear function in two parts, with a knot placed at £40,000. The intention of using a piecewise linear function is to allow for the heterogenous associations by income across the distribution, discovered in the non-parametric analysis and discussed above. This model places into the regression framework the raw gap in university admissions by household income, also allowing a baseline against which the other models may be compared.

In the second model (M2) I additionally control for prior attainment at KS2. Human capital theory predicts that people will invest in education up to the point at which it no longer has a marginal net present benefit. It also implies that those with greater ability gain greater benefits from investing in more education and face lower costs in doing so. As such, as ability increases we would expect the probability of application to, and attendance at, HE to increase also. However, ability itself is not directly observable, instead we must use proxies such as prior attainment. In the LSYPE the earliest measure of attainment provided is KS2 results. Unfortunately, ability is not the only factor correlated with educational attainment. In particular, other socioeconomic factors such as parental education and household income are likely also to be covariates. Given this importance of socioeconomic factors to success in primary education, models using the KS2 attainment data in the LSYPE will recover estimates which must be interpreted carefully. Simple interpretation of coefficients on attainment variables as if they measured true ability would overstate the impact of this underlying factor. Likewise, coefficients on household income are likely to be downward biased relative to a complete income gradient across the lifespan, because of the presumed impact of socioeconomic background on attainment at KS2.

In the third model (M3) I add socioeconomic factors: month of birth, ethnic group, government office region, number of siblings, number of older siblings, whether family type is lone parent or couple, and parental education. These are primarily measured at Wave 1 (age 14), but data from later waves are used where this was missing. Since most are time invariant I assume that this is not problematic. Including these characteristics attempts to separate out some of the other observable socioeconomic effects which are likely correlated with household income. The estimates correspond to what might be though of as conditional income gradients. Hence the model's estimates should identify the household income gradient for a young person with an otherwise similar socioeconomic status. There is some potential for a subset of these characteristics to be endogenous, in that they are themselves partially determined by household income. For example, low levels of household income could be linked with increased chance of a child being raised in a lone parent household. This would result in an underestimate of the impact of household income.

My fourth model (M4) returns to simply controlling for prior attainment, this time up to KS4, using individuals' capped GCSE point scores. As with KS2 attainment this controls for a proxy of ability, but will absorb some of the impact of socioeconomic variables, due to their impact on GCSE attainment. It does, however, allow us some insight into the remaining income gradients that may or may not persist once attainment at 16 is held constant. I leave the KS2 attainment in the model as a safeguard. In the fifth model (M5), I once again add socioeconomic factors. This time attempting to separate the household income gradient from observable confounding factors, holding prior attainment constant at age 16 - that is the conditional income gradient after this point.

In the sixth (M6) and seventh (M7) models I use two techniques to control additionally for the effects a young person's secondary school is likely to exert on HE application and participation. While this allows us to draw additional inference about school effects on university application and attendance, once again the remaining effects attributed to socio-economic characteristics, including household income, may be an underestimate. This is because a young person's socioeconomic characteristics help to determine the secondary school that they attend. The most extreme example of this will be independent schools: an individual's household income is highly correlated with their probability of attending this school type. In this regard, some portion of the impact on university application and attendance can be seen as being a function of household income and other socioeconomic characteristics. For both of these models I include prior attainment at KS2, not KS4, since GCSE results will likely already reflect much of the impact due to Key Stage 3 schooling characteristics.

M6 uses dummy variables for school type characteristics. School type characteristics included were whether the school is a community school, a community technology college, a foundation school, an independent school, a voluntary aided school or a voluntary controlled school. Additionally, dummy variables were included indicating whether the school is a grammar school and whether is has a sixth form. This should allow identification of the impact of specific school characteristics on university admissions. However, it leaves more potential for omitted variable bias where other observed or unobserved school characteristics haven't been included in the model.

$$Pr(\text{Apply or Attend})_{is} = \alpha_0 + \beta\text{Income≤£40,000} + \beta\text{Income>£40,000} + \boldsymbol{\gamma}\mathbf{X} + \zeta_s + \varepsilon_{is} \quad (4)$$

For M7, I allow for school fixed effects. This uses the survey design (sampling was clustered on schools) to remove variation associated with the school an individual is in. I estimate a linear probability regression model[3] of the form shown in equation 4, where α_0 represents the probability for a base set of characteristics; **X** represents a vector of additional controls; ζ_s is a school specific error term; and ε_{is} is an individual error term.

Allowing for fixed effects has drawbacks. It prevents inference from being drawn about specific school characteristics; however, given the main focus of this paper is not on the effects of schooling characteristics I am happy to make this sacrifice. On the positive side, fixed effects offers a more robust method of eliminating bias caused by correlation between household income and unobserved schooling characteristics from estimates. In fact, M6 and M7 tend to produce reasonably similar results.

Estimates of the household income gradients in the various models are in Table 5, while a summary of the results demonstrated through predicted probabilities for each household income quintile is given for females in Table 6 and for males in Table 7. Full reporting of the marginal effects of the characteristics at sample means estimated in the models is given in Appendix B.

A £10,000 change in household equivalised income is equivalent to a family of four's income increasing by £20,000. This is, for example, equivalent to that family's position in the income distribution shifting from the 36th percentile to the 74th percentile. It is a shift of 0.7 of a standard deviation.

Considering first the attendance models, predicted probabilities for income quintiles suggest that females (males) in the top quintile group are roughly 47 percentage points or 3 times (49 percentage point or 3.6 times) more likely to attend university than those in the bottom quintile group. Whichever way it is looked at, this seems a large gap.[4] Comparing

[3]The fixed effects probit estimator is inconsistent, hence I use a linear probability model for convenience.

[4]One caveat that should be noted is that this is somewhat larger than comparison of quintiles in the raw

Table 5: Estimates of university application and admissions marginal effects associated with a £10,000 change in equivalised household permanent income

Var.	M1	M2	M3	M4	M5	M6	M7
Attend - Female: N = 4036							
Income $\leq$ £40k	0.142 (0.009)***	0.070 (0.009)***	0.050 (0.010)***	0.019 (0.007)**	0.027 (0.008)***	0.039 (0.010)***	0.032 (0.010)***
Income $>$ £40k	0.044 (0.023)*	0.048 (0.024)**	0.038 (0.022)*	0.020 (0.021)	0.013 (0.017)	0.011 (0.019)	0.014 (0.014)
Attend - Male: N = 3812							
Income $\leq$ £40k	0.135 (0.009)***	0.076 (0.008)***	0.052 (0.009)***	0.017 (0.007)**	0.018 (0.008)**	0.040 (0.010)***	0.035 (0.010)***
Income $>$ £40k	0.013 (0.019)	0.019 (0.016)	0.008 (0.012)	0.012 (0.011)	0.005 (0.009)	-0.005 (0.010)	-0.006 (0.013)
Apply - Female: N = 4036							
Income $\leq$ £40k	0.156 (0.010)***	0.080 (0.010)***	0.065 (0.011)***	0.022 (0.008)***	0.037 (0.009)***	0.052 (0.011)***	0.043 (0.010)***
Income $>$ £40k	0.050 (0.023)**	0.062 (0.027)**	0.052 (0.023)**	0.047 (0.023)**	0.029 (0.017)*	0.014 (0.016)	0.008 (0.011)
Apply - Male: N = 3812							
Income $\leq$ £40k	0.147 (0.010)***	0.086 (0.009)***	0.055 (0.010)***	0.019 (0.007)**	0.016 (0.009)*	0.039 (0.010)***	0.034 (0.010)***
Income $>$ £40k	-0.003 (0.019)	0.005 (0.017)	-0.011 (0.012)	-0.004 (0.012)	-0.015 (0.011)	-0.028 (0.012)	-0.023 (0.013)
Conditional Attend - Female: N = 2633							
Income $\leq$ £40k	0.056 (0.011)***	0.025 (0.010)**	0.014 (0.012)	0.005 (0.009)	0.006 (0.011)	0.010 (0.011)	0.001 (0.011)
Income $>$ £40k	0.027 (0.024)	0.028 (0.023)	0.021 (0.022)	0.015 (0.021)	0.011 (0.020)	0.008 (0.021)	0.010 (0.012)
Conditional Attend - Male: N = 2204							
Income $\leq$ £40k	0.065 (0.014)***	0.026 (0.012)**	0.022 (0.014)	0.004 (0.012)	0.009 (0.013)	0.020 (0.014)	0.017 (0.013)
Income $>$ £40k	0.094 (0.040)**	0.092 (0.038)**	0.071 (0.036)**	0.067 (0.032)**	0.052 (0.029)*	0.060 (0.035)*	0.011 (0.013)
Socioecon.			√		√	√	√
KS2 Attain		√	√	√	√	√	√
KS4 Attain				√	√		
School						Char.	FE

Notes: Marginal effects estimated at sample means. Adjusted using LSYPE Wave 7 respondent weights, which attempt to adjust for oversampling and attrition. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised permanent income. Sample: Wave 7 Participants with valid responses for all variables used in models.

Table 6: Predicted Probabilities by income Quintile group - Female

Variable	Q1	Q2	Q3	Q4	Q5	N
Model 1 - Female						
University Application	0.35	0.44	0.52	0.63	0.84	4036
University Attend	0.24	0.32	0.39	0.50	0.73	4036
University Attend (Conditional on Apply)	0.69	0.73	0.75	0.79	0.87	2633
Model 2 - Female						
University Application	0.39	0.45	0.49	0.56	0.73	4036
University Attend	0.27	0.32	0.36	0.42	0.58	4036
University Attend (Conditional on Apply)	0.69	0.71	0.72	0.74	0.80	2633
Model 3 - Female						
University Application	0.36	0.41	0.45	0.51	0.67	4036
University Attend	0.29	0.33	0.36	0.41	0.54	4036
University Attend (Conditional on Apply)	0.76	0.77	0.78	0.79	0.83	2633
Model 4 - Female						
University Application	0.32	0.34	0.35	0.38	0.47	4036
University Attend	0.18	0.19	0.20	0.22	0.27	4036
University Attend (Conditional on Apply)	0.55	0.55	0.55	0.56	0.59	2633
Model 5 - Female						
University Application	0.21	0.24	0.26	0.30	0.42	4036
University Attend	0.14	0.16	0.17	0.20	0.26	4036
University Attend (Conditional on Apply)	0.55	0.56	0.56	0.57	0.60	2633
Model 6 - Female						
University Application	0.36	0.40	0.44	0.49	0.61	4036
University Attend	0.29	0.32	0.34	0.38	0.47	4036
University Attend (Conditional on Apply)	0.76	0.77	0.77	0.78	0.80	2633
Model 7 - Female						
University Application	0.45	0.47	0.49	0.53	0.60	4036
University Attend	0.36	0.38	0.39	0.42	0.47	4036
University Attend (Conditional on Apply)	0.78	0.78	0.78	0.78	0.79	2633

Notes: Other model characteristics held fixed as follows: White British ethnicity, born in November, from London, eldest child of two, both parents with GCSEs or equivalent as highest level of education, average raw score of 60 at Key Stage 2 SATS, capped GCSE score of 306 and attended a non-selective Community School (with a Sixth Form) for Key Stage 3. Adjusted using LSYPE Wave 7 respondent weights.

Table 7: Predicted Probabilities by income quintile group - Male

Variable	Q1	Q2	Q3	Q4	Q5	N
Model 1 - Male						
University Application	0.28	0.36	0.43	0.54	0.74	3812
University Attend	0.18	0.24	0.31	0.41	0.65	3812
University Attend (Conditional on Apply)	0.64	0.68	0.71	0.76	0.87	2204
Model 2 - Male						
University Application	0.28	0.34	0.39	0.46	0.61	3812
University Attend	0.17	0.21	0.25	0.31	0.47	3812
University Attend (Conditional on Apply)	0.62	0.64	0.65	0.68	0.78	2204
Model 3 - Male						
University Application	0.28	0.32	0.35	0.40	0.50	3812
University Attend	0.16	0.19	0.22	0.26	0.37	3812
University Attend (Conditional on Apply)	0.62	0.64	0.65	0.67	0.76	2204
Model 4 - Male						
University Application	0.30	0.31	0.33	0.35	0.38	3812
University Attend	0.16	0.18	0.19	0.20	0.25	3812
University Attend (Conditional on Apply)	0.56	0.56	0.56	0.57	0.63	2204
Model 5 - Male						
University Application	0.21	0.22	0.23	0.25	0.27	3812
University Attend	0.10	0.11	0.11	0.13	0.16	3812
University Attend (Conditional on Apply)	0.48	0.49	0.49	0.50	0.58	2204
Model 6 - Male						
University Application	0.31	0.34	0.37	0.40	0.46	3812
University Attend	0.16	0.19	0.21	0.24	0.31	3812
University Attend (Conditional on Apply)	0.57	0.59	0.61	0.63	0.71	2204
Model 7 - Male						
University Application	0.31	0.33	0.35	0.38	0.41	3812
University Attend	0.17	0.20	0.21	0.24	0.29	3812
University Attend (Conditional on Apply)	0.61	0.62	0.63	0.65	0.68	2204

Notes: Other model characteristics held fixed as follows: White British ethnicity, born in November, from London, eldest child of two, both parents with GCSEs or equivalent as highest level of education, average raw score of 60 at Key Stage 2 SATS, capped GCSE score of 306 and attended a non-selective Community School (with a Sixth Form) for Key Stage 3. Adjusted using LSYPE Wave 7 respondent weights.

this with the Apply models, we find that the gap between the top and bottom quintiles is similar at 49 (46) percentage points for females (males). It comes as little surprise then that our first Conditional Attend model identifies a relatively small (but significant) association, even with no controlling factors: those in the top quintile are just 18 (23) percentage points for females (males) more likely to get into university, having applied.

As one would expect, and previous research suggests, these associations are much reduced once additional covariates are controlled for. The base regression model takes no account of prior attainment, which acts both as an imperfect measure of underlying ability and as a function of socioeconomic characteristics on attainment up to that point. Once KS2 attainment is included in M2 the attendance gap between top and bottom quintiles falls to 31 (30) percentage points. The relatively small effect of income on attendance, conditional on having applied, becomes even smaller, in this case the quintile gap closes to 11 (16) percentage points for females (males). Although smaller, this is perhaps a more robust indication of socioeconomic variation in university attendance, once a correlate of ability has been controlled for.

We observe further drops once socioeconomic characteristics are added in M3 and the marginal effects for conditional attendance become insignificant. However, also interesting given the specification is to examine the other significant associations. There are significant marginal effects for the ethnicity dummy variables, sibling effect dummy variables and a small but significant effect of month of birth. Notably, we see significant estimated effects for lone parent family status and some parental education variables. Father having a degree relative to holding GCSE qualifications, in particular, shows a large and significant marginal effect comparable to £20-30,000 of additional household equivalised income.

For M4, we return to just controlling for prior attainment, this time up to Key Stage 4/GCSE. For attendance the gap between the top and bottom income quintiles drops to 9 percentage points. This is in contrast to much previous research where no significant effect of socioeconomic status is generally identified once educational outcomes at the age of 16 are controlled for (Marcenaro-Gutierrez et al., 2007, p.351). These results suggest that there is

data, which suggest a gap of 42 percentage points or 2.75 times. Imposing a parametric structure has perhaps exaggerated the gap somewhat. Interrogation of the predicted probability and non-parametric graphs suggest this is related to the inverted relationship at the bottom of the income distribution.

a small, but significant, effect of household income after the age of 16, but much of the association from earlier models was having effect through improved educational performance earlier in the school career. We can again use the application and conditional attendance models to see that this effect seems to be driven by the application decision: for conditional attendance the remaining gaps are small and not statistically significant.

Very little change in the apparent effect of income is observed when socioeconomic factors are also included in M5. Nevertheless, some of the the same parental education variables once again show a smaller, but still significant marginal effect.

In M7, the gap between top and bottom income quintile groups is 11 percentage points. It seems surprising that even within a school and for individuals with otherwise very similar socioeconomic characteristics a small association between household income and university attendance is still identified.

In the conditional attendance models generally there are very few statistically significant coefficients. It is interesting to examine what these are. In all relevant models, the coefficients for prior attainment are jointly significant [5]. In M6, for females, in addition to prior attainment the model identifies a positive significant effect on attending either an independent or grammar school and father having education to degree level. It also identifies a negative significant effect for mother having no educational qualifications and, somewhat anomalously, father's highest level of education being A-Level or equivalents. We should not perhaps be surprised by the small and occasionally unexpected remaining coefficients on parental education. Parental education would seem to be more important for a young person's achievement earlier in their educational career with, perhaps, only highly educated parents able to provide additional help to their offspring throughout secondary school.

Checking this for robustness by using the fixed effect model (M7), we cannot, of course, identify effects due to individual school types, but father holding a degree remains significant, while other parental education effects become insignificant. For males, the picture is slightly different. In M6, other than prior attainment (again jointly significant), the only significant effects identified are for attending a grammar school and father holding a de-

[5] Prior attainment is modelled using a quadratic and/or piecewise function, and although individual coefficients may not be significant a Wald test of joint significance always rejects the null hypothesis of no impact.

gree. Overall, the picture is of very little other than prior attainment playing a role in the probability of attendance conditional on having applied.

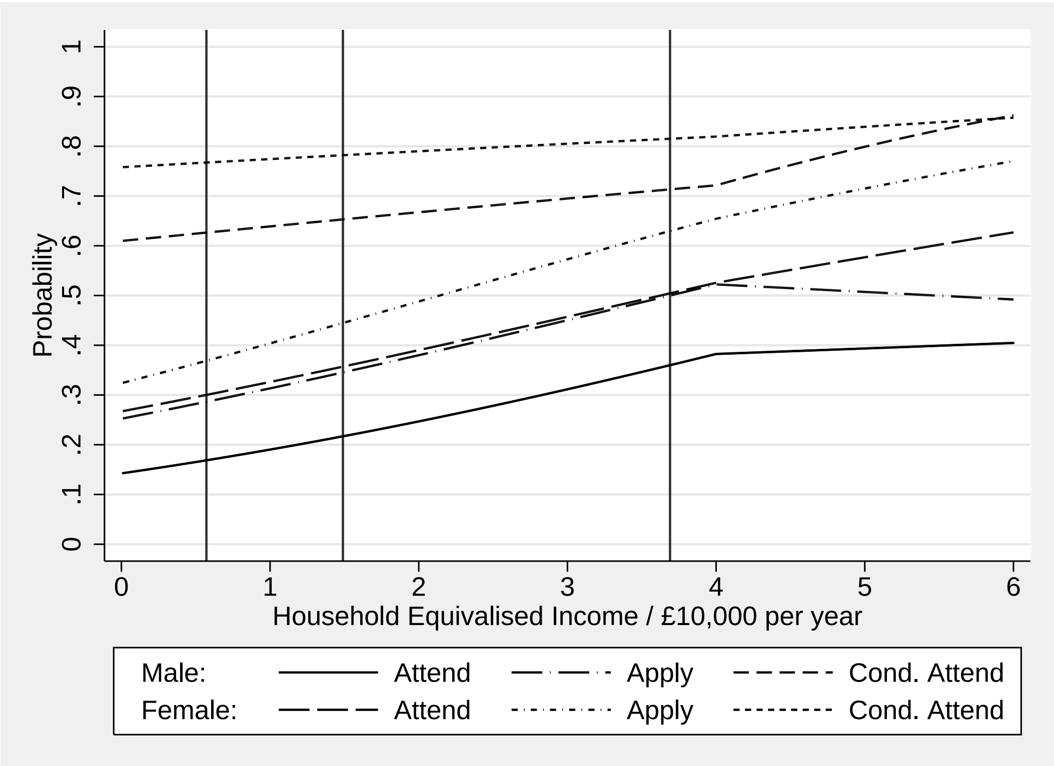
Graphs showing the predicted probabilities of the relevant outcomes across the income range help to place the findings in their context. These require a fixed set of other characteristics in the regression model. In order to make this as generally representative as possible I have generally chosen the most commonly occurring value of each categorical variable and the mean level of prior attainment variables (rounded to the nearest whole number). As such the graphs show predicted probabilities for those with the following characteristics: White British ethnicity, born in November, located in London, eldest child of two, both parents with GCSEs or equivalent as highest level of education, average raw score of 60 at KS2 SATS, average GCSE capped point score of 306 and attended a non-selective community school with a sixth form for Key Stage 3. The graph in Figure 5 shows the predicted probabilities for Model 3 (controlling for household income, KS2 prior attainment and socioeconomic background) as a fairly typical example.

A further observation from these is the differences between the probabilities by gender. In both Apply and Attend models the probabilities for females are generally higher, as shown by the upward shift of the line relative to men. This is in line with the simple difference in proportions applying and attending by gender. In addition, the lines of probabilities also have a slightly steeper gradient. This may be interpreted as an increased importance of the role of household income in the application decision for females. As remarked earlier, the statistical significance of this gradient difference is in doubt and so should be interpreted as suggestive only.

However, Figure 6, showing predictions from Model 4, is notable for the absence of that gap between males and females. Given that Model 4 includes Key Stage 4/GCSE attainment this gives an indication that a great deal of the gender difference in these rates is driven by difference in attainment by this stage.

The top ends of the graphs also prompt me to consider the top part of the income distribution. We gain a clearer picture of the overall form of the relationship between household income and university application once other factors have been controlled for. In the case of females, it would appear that we cannot reject the possibility of a simple linear relationship across the whole income range. Indeed, in a formal Wald test we cannot reject this as

Figure 5: Predicted Probability of University Application, Attendance and Conditional Attendance as Household Equivalised Income varies



Notes: Predicted probabilities from M3 (household income, KS2 prior attainment and socioeconomic characteristics). Other model characteristics held fixed as follows: White British ethnicity; born in November; from London; eldest child of two; both parents with GCSEs or equivalent as highest level of education; and average raw score of 60 at Key Stage 2 SATS. Sample size: 7850. Vertical lines show 10th, 50th and 90th percentiles of income.

Figure 6: Predicted Probability of University Application and Attendance as Household Equivalised Income varies

Notes: Predicted probabilities from M4 (household income, KS2 prior attainment and KS4 prior attainment). Other model characteristics held fixed as follows: Average raw score of 60 at Key Stage 2 SATS and average GCSE capped point score of 306. Sample size: 7850. Vertical lines show 10th, 50th and 90th percentiles of income.

a null hypothesis even at the 10% level. On the contrary, for males, we can reject this null hypothesis at the 1% level. This also holds for university attendance.

The finding of small and often insignificant gradients for household income conditional on having applied is reassuring, if we assume that otherwise similar individuals should not be advantaged or disadvantaged in the admissions process by their household income. However, a key question is left unanswered. Although individuals with different household incomes seem to stand a similar chance of getting into university, so long as they apply, do they get into similar universities? The above analysis remains silent on this matter, but it is what I now turn to.

7 Comparison between Russell Group and others

The Russell Group refers to a group of twenty research intensive UK universities[6]. They are often considered to be the most prestigious UK universities. Since they are an unusual group of universities we might expect the determinants of attending a Russell Group university to be different from universities in general. In particular, there would seem to be the possibility that although we saw only small associations between income and achieving a place at university overall, those with high levels of income could be disproportionately attending high quality institutions.

Previous research has considered differences by school type in the patterns of applications to the 'Sutton 13'[7]. Using UCAS data Department for Business, Innovation & Skills (2009) presented evidence that, for a given level of attainment, those who applied to a 'Sutton 13' university were not more or less likely to receive an offer dependent on their school type. However, the probability of application to a 'Sutton 13' institution did vary by school type,

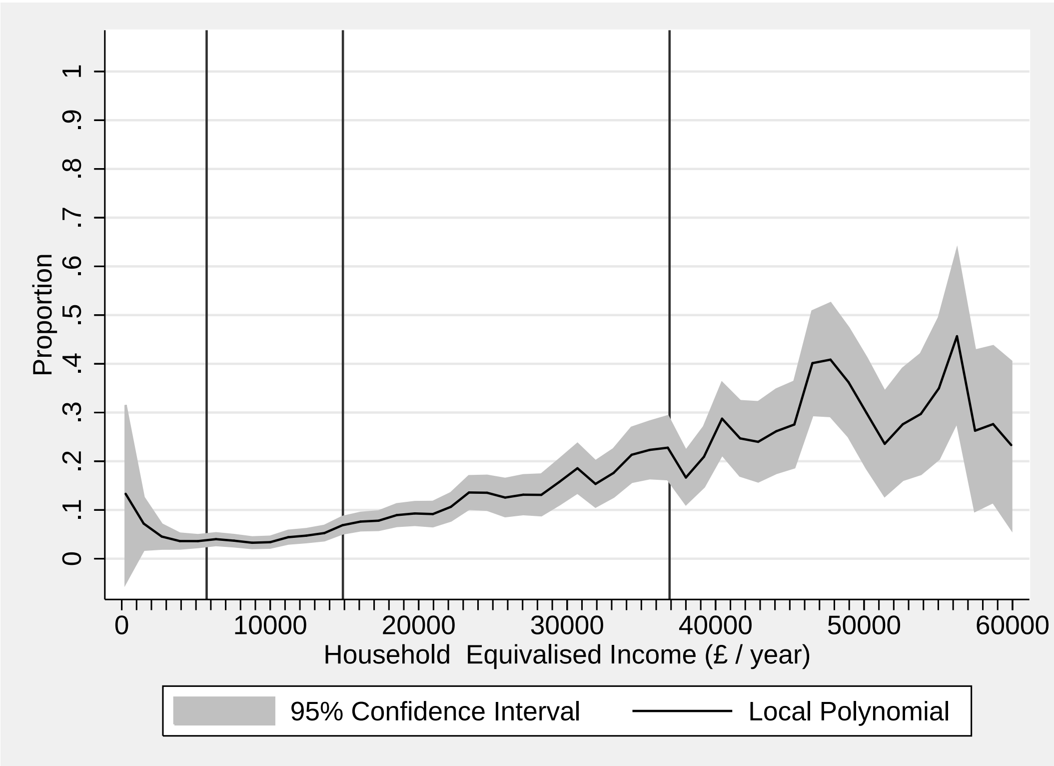
[6] As of March 2012 four additional institutions have joined the Russell Group. However, given the timeframe of the data collection, for our purposes the Russell Group is made up of the following twenty universities: University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, University of Edinburgh, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Liverpool, London School of Economics and Political Science, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, Queen's University Belfast, University of Sheffield, University of Southampton, University College London and University of Warwick.

[7] The 'Sutton 13' is an alternative grouping of 'elite' universities drawn up by the Sutton Trust. It includes the following institutions: University of Birmingham, University of Bristol, University of Cambridge, Durham University, University of Edinburgh, Imperial College, London School of Economics, University of Nottingham, University of Oxford, University of St Andrews, University College London, University of Warwick and University of York.

even after conditioning on average attainment within schools. This research did not have the rich socioeconomic background data available in the LSYPE.

The proportion of the LSYPE cohort who attend a Russell Group university is 9%, while the proportion of those who attend university that attend a Russell Group university is 25%. Another comparison worth drawing here is that while 76% of those who apply to university get into one, only just under 19% of those who apply to university get into a Russell Group institution. Our truly comparable measure is missing here, since we do not observe whether individuals apply to a Russell Group university or not. As we know, it is not the case that everyone who applies to university gets a place, but even if we generalised this as true the same cannot be said for the Russell Group. Under these different conditions it could be the case that additional household income can make more of a difference to an individual's chances.

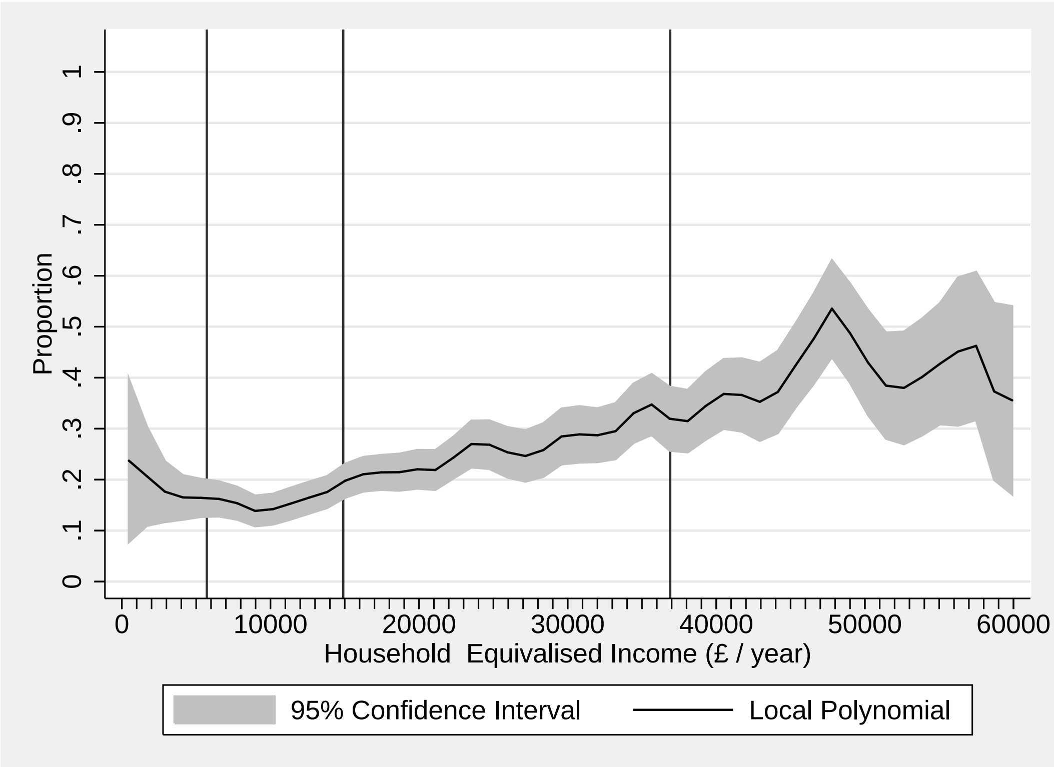
Figure 7: Non-parametric plot of Household Equivalised Income and Russell Group University Attendance



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 1577.335. Sample size: 7780. Vertical lines show 10th, 50th and 90th percentiles of income.

Figure 8 shows the association between household income and attendance at a Russell Group university for those of the sample who attend a university. I again use local poly-

Figure 8: Non-parametric plot of Household Equivalised Income and Russell Group University Attendance, conditional on attending any university



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 2485.145. Sample size: 3771. Vertical lines show 10th, 50th and 90th percentiles of income.

Table 8: Probability of Russell Group university attendance by income quintile group

Variable	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**N**
Russell Group Attendance	0.04	0.04	0.07	0.11	0.24	7849
Russell Group Attendance (Conditional on Uni)	0.17	0.15	0.20	0.24	0.36	3827

Notes: Adjusted using LSYPE Wave 7 respondent weights, which attempt to adjust for oversampling and attrition. Sample: Wave 7 Participants with valid responses for variables used in models.

nomial smoothing, as discussed in Section 5. The upward slope across much of the income range suggests that those from households with higher incomes are more likely to attend Russell Group universities and we see this confirmed in Table 8 using analysis of rates among household income quintiles: those in the top quintile are 19 percentage points more likely to be in a Russell Group university than those in the bottom quintile.

Regression models identical to those for the main analysis were specified, estimating the probability of attending a) a Russell Group university and b) a Russell Group university conditional on attending university. In effect, I am placing an additional layer in the sequential model of university admissions, as described in Section 4. Separate models have not been specified for males and females. I took this decision because the models conditional on university attendance would have rather low sample sizes. Household income gradients are shown in Table 9. As with the main analysis, complete reporting of marginal effects at sample means is included in Appendix B.

Table 9: Estimates of Russell Group attendance marginal effects associated with a £10,000 change in equivalised household permanent income

Var.	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**
Russell Group Attendance: N = 7847							
Income	0.047	0.028	0.017	0.012	0.009	0.013	0.014
$\leq$ £40k	(0.003)***	(0.002)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.005)***
Income	0.004	0.004	0.002	-0.003	-0.003	-0.003	-0.003
$>$ £40k	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.010)
Russell Group Conditional Attendance: N = 3826							
Income	0.063	0.042	0.025	0.024	0.018	0.019	0.012
$\leq$ £40k	(0.006)***	(0.007)***	(0.009)***	(0.007)***	(0.009)**	(0.009)**	(0.010)
Income	0.005	0.001	-0.001	-0.009	-0.008	-0.007	-0.007
$>$ £40k	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.013)
Socioecon.			√		√	√	√
KS2 Attain		√	√	√	√	√	√
KS4 Attain				√	√		
School						Char.	FE

Notes: Marginal effects estimated at sample means. Adjusted using LSYPE Wave 7 respondent weights, which attempt to adjust for oversampling and attrition. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised permanent income. Sample: Wave 7 Participants with valid responses for all variables used in models.

Considering the unconditional model of whether an individual attends a Russell Group university we see that the top income quintile have approximately 23 percentage points higher probability of attending a Russell Group university than those in the bottom income quintile. However, we know from our main analysis that the attendance gap can mask variation

Table 10: Predicted Probabilities by income quintile group - Russell Group

Variable	Q1	Q2	Q3	Q4	Q5	N
Model 1						
Russell Group Attend	0.03	0.05	0.07	0.11	0.26	7847
Russell Group Attend (Conditional on Uni)	0.14	0.17	0.20	0.25	0.37	3826
Model 2						
Russell Group Attend	0.02	0.03	0.04	0.05	0.11	7847
Russell Group Attend (Conditional on Uni)	0.08	0.09	0.10	0.13	0.18	3826
Model 3						
Russell Group Attend	0.01	0.02	0.02	0.03	0.05	7847
Russell Group Attend (Conditional on Uni)	0.05	0.06	0.06	0.07	0.09	3826
Model 4						
Russell Group Attend	0.07	0.08	0.09	0.10	0.13	7847
Russell Group Attend (Conditional on Uni)	0.10	0.12	0.13	0.14	0.17	3826
Model 5						
Russell Group Attend	0.03	0.03	0.04	0.04	0.05	7847
Russell Group Attend (Conditional on Uni)	0.05	0.06	0.06	0.07	0.09	3826
Model 6						
Russell Group Attend	0.01	0.02	0.02	0.02	0.03	7847
Russell Group Attend (Conditional on Uni)	0.05	0.05	0.05	0.06	0.07	3826
Model 7						
Russell Group Attend	0.03	0.04	0.04	0.05	0.07	7847
Russell Group Attend (Conditional on Uni)	0.06	0.07	0.08	0.09	0.10	3826

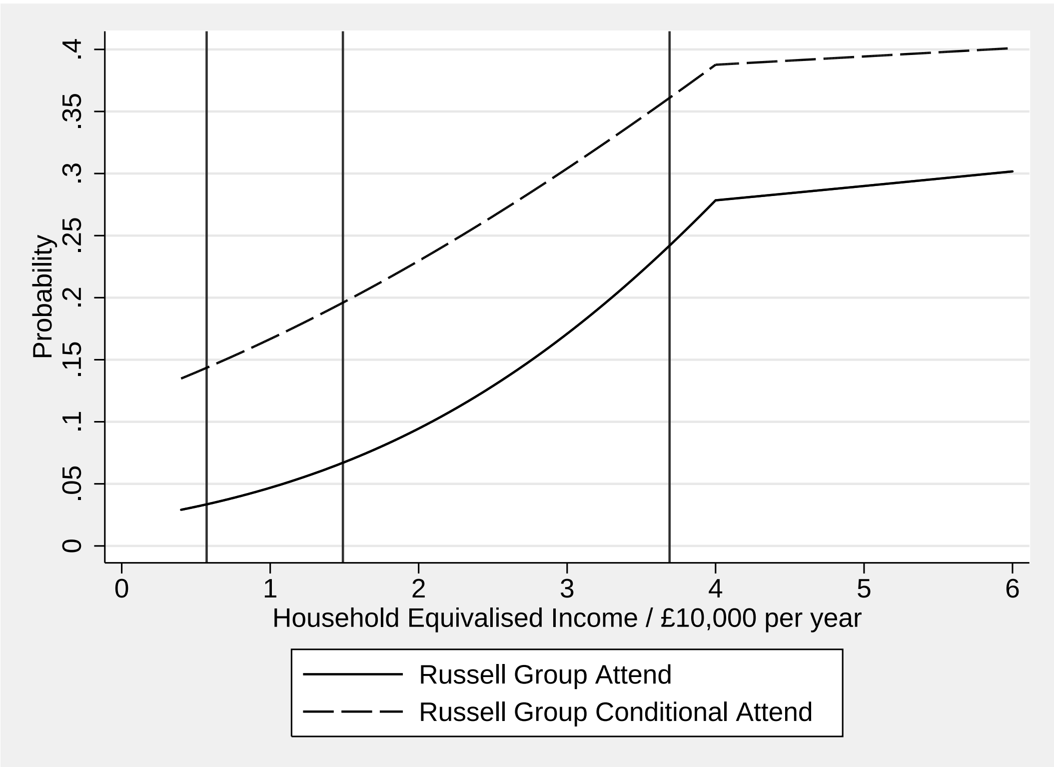
Notes: Other model characteristics held fixed as follows: White British ethnicity, born in November, from London, eldest child of two, both parents with GCSEs or equivalent as highest level of education, average raw score of 69 at Key Stage 2 SATS, capped GCSE score of 381 and attended a non-selective Community School (with a Sixth Form) for Key Stage 3. Adjusted using LSYPE Wave 7 respondent weights.

in how it emerges.

When we consider a model of Russell Group attendance conditional on attending any university (also see graphically in Figure 9) we see that a more pronounced gap by income remains. The gap between the bottom and top income quintiles remains 23 percentage points.

A noticeable feature of the relationship, in comparison with university conditional attendance, is that although it initially drops dramatically with Key Stage 2 prior attainment it does not drop as much with Key Stage 4 prior attainment. In M2 the gap between top and bottom income quintiles more than halves to 10 percentage points, when other socio-economic controls it drops to 4 percentage points but does not fall much further except in the fixed effects model.

Figure 9: Predicted Probability of Russell Group Attendance, conditional on University attendance, as Household Equivalised Income varies



Notes: Predicted probabilities from M1 (household income only). Sample size: 3826. Vertical lines show 10th, 50th and 90th percentiles of income.

At first sight this is quite different from the gap seen for the university conditional attendance models: it cannot be explained away by understandably important factors such as prior attainment. However, this is not a fair comparison: we do not observe the universities individuals have applied to, meaning that we cannot be sure how much of this gap emerges because of the differing application choices of individuals across the household income distribution. An individual cannot, after all, attend a Russell Group university unless he or she applied to one or more of them. The findings from Department for Business, Innovation & Skills (2009) suggest this could well be the case.

8 Conclusions

In some fundamental ways these findings accord with previous research, for example in showing that socioeconomic status is important in predicting whether an individual will attend university. However, this paper goes beyond previous research in several important ways. In the introduction three key tasks were set out for this paper and I believe these have been dealt with.

First, I have quantified the relationship between permanent household income and university attendance for a recent English cohort. These data suggest that those in the top fifth of the income distribution are around 2.75 times as likely to attend university as those in the bottom fifth. This relationship persists, albeit smaller, even once we control for a range of other confounding factors, including some that seem likely to lead to an underestimate of the effect of income. However, even for the model with no controls, this relationship appears much weaker for equivalised household income above £40,000. This may suggest some role for credit constraints in explaining the link.

Second, by analysing the probability of application and the probability of attendance conditional on having applied separately, I demonstrate that the link is predominantly driven by the application decision. Even after controlling for prior attainment and socioeconomic background a significant application gap remains. On the contrary, I identify only a small household income gradient for attendance conditional on having applied and show that, conditional on having applied, those in the top fifth of the income distribution are approximately 1.25 times more likely to attend than those in the bottom fifth. Even this rapidly vanishes once additional controls such as earlier educational outcomes are added.

Finally, I analysed attendance at Russell Group universities, as a special case of high quality institutions. There does, here, appear to be a relatively small but resilient gap in attendance at high status universities by household income. Without better data on the application choices of individuals it is impossible to shed light on whether this is similarly driven by those application choices or emerges during the admissions process.

A key message of this paper is that the university participation gap largely emerges at or before young people apply. This suggests that narrowing the gap through policy intervention at the point of admissions may be very difficult, since in the population such policies would work on there is not so much gap left. The policies could only have a significant effect if they also led to a change in perception of the university application process, in turn leading to a broader application population. More likely to be successful are policies which intervene earlier to ensure that those from poorer backgrounds achieve their potential during their educational career and hence are more likely to have the appropriate qualifications to apply to university.

Bibliography

Acemoglu, D. and Pischke, J.-S. (2001). Changes in the wage structure, family income, and children's education. *European Economic Review*, 45(4-6):890–904.

Anders, J. (2012). Using the Longitudinal Study of Young People in England (LSYPE)/"Next Steps" for research into Higher Education access. mimeo., forthcoming.

Anders, J., Dearden, L., and Sianesi, B. (2012). The Role of Credit Constraints in Educational Choices: Evidence from the LSYPE. mimeo., forthcoming.

Atkinson, A. B. (1983). *The Economics of Inequality*. Clarendon Press, Oxford.

Becker, G. S. and Tomes, N. (1986). Human-Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4(3):S1–S39.

Bekhradnia, B. (2003). Widening participation and fair access: An overview of the evidence. Report, Higher Education Policy Institute.

Blanden, J. and Gregg, P. (2004). Family income and educational attainment: a review of approaches and evidence for Britain. CEE Working Papers 0041, Centre for the Economics of Education, London School of Economics, London.

Blanden, J. and Machin, S. (2004). Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy*, 51(2):230–249.

Blau, D. M. (1999). The effect of income on child development. *The Review of Economics and Statistics*, 81(2):261–276.

Breen, R. and Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society*, 9(3):275–305.

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

Broecke, S. and Hamed, J. (2008). Gender gaps in higher education participation: an analysis of the relationship between prior attainment and young participation by gender, socio-economic class and ethnicity. DIUS Research Report 08 14, Department for Innovation, Universities and Skills.

Cameron, D. (2010). PM's speech on education. `http://www.number10.gov.uk/news/pms-speech-on-education/`.

Carneiro, P. and Heckman, J. J. (2002). The Evidence on Credit Constraints in Post-Secondary Schooling. *The Economic Journal*, 112(482):705–734.

Chevalier, A. and Conlon, G. (2003). Does It Pay to Attend a Prestigious University? CEE Discussion Paper 40, Centre for the Economics of Education, London School of Economics.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2010). Widening participation in higher education: analysis using linked administrative data. IFS Working Paper W10/04, Institute of Fiscal Studies.

Dearden, L., McGranahan, L., and Sianesi, B. (2004). The Role of Credit Constraints in Educational Choices: Evidence from NCDS and BCS70. CEE Discussion Paper, Centre for the Economics of Education, London School of Economics.

Department for Business, Innovation & Skills (2009). Applications, Offers and Admissions to Research Led Universities. BIS Research Paper 5, Department of Business, Innovation & Skills.

Department for Business, Innovation & Skills (2011). Participation Rates in Higher Education: Academic Years 2006/2007–2009/2010 (Provisional). Statistical first release, Department for Business, Innovation & Skills.

Department for Education and National Centre for Social Research (2011). Longitudinal Study of Young People in England: Waves One to Seven, 2004-2010 [computer file].

Gayle, V., Berridge, D., and Davies, R. (2000). Young people's routes to higher education: exploring processes with longitudinal data. *Higher Education Review*, 33(1):52–63.

Gayle, V., Berridge, D., and Davies, R. B. (2003). Econometric Analysis of the Demand for Higher Education. DfES Research Reports RR472, Department for Education and Skills.

Haveman, R. and Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, 33(4):1829–1878.

HM Government (2011). Opening Doors, Breaking Barriers: A Strategy for Social Mobility.

Hobbs, G. and Vignoles, A. (2010). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36(4).

Jenkins, S. P. and Schluter, C. (2002). The Effect of Family Income During Childhood on Later-Life Attainment: Evidence from Germany. IZA Discussion Papers 604, Institute for the Study of Labor (IZA).

Marcenaro-Gutierrez, O., Galindo-Rueda, F., and Vignoles, A. (2007). Who actually goes to university? *Empirical Economics*, 32(2):333–357.

Mayer, S. E. (2002). The Influence of Parental Income on Children's Outcomes. Research report, New Zealand Ministry of Social Development.

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics*, 77(2):155–184.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Thomas, R. L. (2005). *Using Statistics in Economics*. McGraw-Hill Education.

Todd, P. and Wolpin, K. I. (2004). The Production of Cognitive Achievement in Children:

Home, School and Racial Test Score Gaps. PIER Working Paper 04-019, Penn Institute for Economic Research, University of Pennsylvania.

A Full summary statistics tables

Table 11: Cross tabulation of university application and attendance

University Applicant	**University Attendee**		
	No	**Yes**	**Total**
	Cell %	Cell %	Cell %
No	50.7	0.0	50.7
Yes	10.7	38.6	49.3
Total	61.4	38.6	100.0

Table 12: Summary Statistics by Sex

Variable	Mean Female	Mean Male	Mean Overall	Standard Deviation
Male	0.00	1.00	0.49	0.50
	(0.00)	(0.00)	(0.01)	
University Applicant	0.55	0.46	0.51	0.50
	(0.01)	(0.01)	(0.01)	
University Attendee	0.43	0.35	0.39	0.49
	(0.01)	(0.01)	(0.01)	
Household 'permanent' income	32155.08	32519.67	32335.21	23608.72
	(572.75)	(547.82)	(427.86)	
Number of Siblings	1.42	1.41	1.42	0.88
	(0.02)	(0.02)	(0.01)	
Number of Older Siblings	0.88	0.85	0.86	0.92
	(0.02)	(0.02)	(0.01)	
Family Type Couple	0.77	0.80	0.78	0.41
	(0.01)	(0.01)	(0.01)	
Mother has L1 or lower quals	0.10	0.09	0.09	0.29
	(0.01)	(0.01)	(0.00)	
Mother has GCSE A-Cs or equiv.	0.32	0.32	0.32	0.47
	(0.01)	(0.01)	(0.01)	
Mother has A Levels or equiv.	0.15	0.15	0.15	0.36
	(0.01)	(0.01)	(0.00)	
Mother has HE below degree	0.14	0.14	0.14	0.34
	(0.01)	(0.01)	(0.01)	
Mother has Degree or equiv.	0.11	0.12	0.12	0.32
	(0.01)	(0.01)	(0.01)	
Father has L1 or lower quals	0.08	0.08	0.08	0.27
	(0.01)	(0.01)	(0.00)	
Father has GCSE A-Cs or equiv.	0.29	0.29	0.29	0.45
	(0.01)	(0.01)	(0.01)	
Father has A Levels or equiv.	0.17	0.19	0.18	0.38
	(0.01)	(0.01)	(0.01)	
Father has HE below degree	0.12	0.12	0.12	0.33
	(0.01)	(0.01)	(0.00)	
Father has Degree or equiv.	0.14	0.14	0.14	0.35
	(0.01)	(0.01)	(0.01)	
KS2 Average Raw Point Score	60.98	60.96	60.98	14.67
	(0.37)	(0.35)	(0.28)	
Community Tech. KS3 College	0.01	0.01	0.01	0.09
	(0.01)	(0.00)	(0.00)	
Community KS3 School	0.65	0.63	0.64	0.48
	(0.02)	(0.02)	(0.02)	
Foundation KS3 School	0.16	0.17	0.16	0.37
	(0.02)	(0.02)	(0.02)	
Independent KS3 School	0.05	0.04	0.05	0.21
	(0.01)	(0.01)	(0.01)	
Voluntary Aided KS3 School	0.11	0.11	0.11	0.31
	(0.01)	(0.01)	(0.01)	
Voluntary Controlled KS3 School	0.02	0.04	0.03	0.17
	(0.01)	(0.01)	(0.01)	
KS3 School is Grammar/Selective	0.03	0.06	0.04	0.20
	(0.01)	(0.01)	(0.00)	
KS3 School has Sixth Form	0.58	0.57	0.58	0.49
	(0.02)	(0.02)	(0.02)	
N	3749	3594	7346	

Notes: Adjusted using LSYPE Wave 7 respondent weights. Standard errors in brackets.

Table 13: Summary Statistics by University Applicant

Variable	Mean Not University Applicant	Mean University Applicant	Mean Overall	Standard Deviation
Male	0.53	0.45	0.49	0.50
	(0.01)	(0.01)	(0.01)	
University Applicant	0.00	1.00	0.51	0.50
	(0.00)	(0.00)	(0.01)	
University Attendee	0.00	0.77	0.39	0.49
	(0.00)	(0.01)	(0.01)	
Household 'permanent' income	25514.68	38953.03	32335.21	23608.72
	(402.13)	(572.18)	(427.86)	
Number of Siblings	1.47	1.37	1.42	0.88
	(0.02)	(0.02)	(0.01)	
Number of Older Siblings	0.95	0.78	0.86	0.92
	(0.02)	(0.02)	(0.01)	
Family Type Couple	0.72	0.84	0.78	0.41
	(0.01)	(0.01)	(0.01)	
Mother has L1 or lower quals	0.13	0.06	0.09	0.29
	(0.01)	(0.00)	(0.00)	
Mother has GCSE A-Cs or equiv.	0.36	0.28	0.32	0.47
	(0.01)	(0.01)	(0.01)	
Mother has A Levels or equiv.	0.14	0.16	0.15	0.36
	(0.01)	(0.01)	(0.00)	
Mother has HE below degree	0.10	0.18	0.14	0.34
	(0.01)	(0.01)	(0.01)	
Mother has Degree or equiv.	0.04	0.19	0.12	0.32
	(0.00)	(0.01)	(0.01)	
Father has L1 or lower quals	0.11	0.05	0.08	0.27
	(0.01)	(0.00)	(0.00)	
Father has GCSE A-Cs or equiv.	0.33	0.25	0.29	0.45
	(0.01)	(0.01)	(0.01)	
Father has A Levels or equiv.	0.18	0.18	0.18	0.38
	(0.01)	(0.01)	(0.01)	
Father has HE below degree	0.09	0.15	0.12	0.33
	(0.01)	(0.01)	(0.00)	
Father has Degree or equiv.	0.05	0.23	0.14	0.35
	(0.00)	(0.01)	(0.01)	
KS2 Average Raw Point Score	54.42	67.34	60.98	14.67
	(0.33)	(0.27)	(0.28)	
Community Tech. KS3 College	0.01	0.01	0.01	0.09
	(0.00)	(0.01)	(0.00)	
Community KS3 School	0.72	0.56	0.64	0.48
	(0.02)	(0.02)	(0.02)	
Foundation KS3 School	0.16	0.17	0.16	0.37
	(0.02)	(0.02)	(0.02)	
Independent KS3 School	0.01	0.08	0.05	0.21
	(0.00)	(0.01)	(0.01)	
Voluntary Aided KS3 School	0.08	0.13	0.11	0.31
	(0.01)	(0.02)	(0.01)	
Voluntary Controlled KS3 School	0.03	0.04	0.03	0.17
	(0.01)	(0.01)	(0.01)	
KS3 School is Grammar/Selective	0.01	0.07	0.04	0.20
	(0.00)	(0.01)	(0.00)	
KS3 School has Sixth Form	0.55	0.60	0.58	0.49
	(0.02)	(0.02)	(0.02)	
N	3618	3728	7346	

Notes: Adjusted using LSYPE Wave 7 respondent weights. Standard errors in brackets.

Table 14: Summary Statistics by University Attendee

Variable	Mean Not University Attendee	Mean University Attendee	Mean Overall	Standard Deviation
Male	0.52	0.44	0.49	0.50
	(0.01)	(0.01)	(0.01)	
University Applicant	0.19	1.00	0.51	0.50
	(0.01)	(0.00)	(0.01)	
University Attendee	0.00	1.00	0.39	0.49
	(0.00)	(0.00)	(0.01)	
Household 'permanent' income	26777.30	41063.50	32335.21	23608.72
	(392.88)	(636.88)	(427.86)	
Number of Siblings	1.46	1.34	1.42	0.88
	(0.02)	(0.02)	(0.01)	
Number of Older Siblings	0.94	0.75	0.86	0.92
	(0.02)	(0.02)	(0.01)	
Family Type Couple	0.74	0.85	0.78	0.41
	(0.01)	(0.01)	(0.01)	
Mother has L1 or lower quals	0.12	0.05	0.09	0.29
	(0.01)	(0.00)	(0.00)	
Mother has GCSE A-Cs or equiv.	0.35	0.27	0.32	0.47
	(0.01)	(0.01)	(0.01)	
Mother has A Levels or equiv.	0.14	0.16	0.15	0.36
	(0.01)	(0.01)	(0.00)	
Mother has HE below degree	0.11	0.19	0.14	0.34
	(0.01)	(0.01)	(0.01)	
Mother has Degree or equiv.	0.06	0.21	0.12	0.32
	(0.00)	(0.01)	(0.01)	
Father has L1 or lower quals	0.10	0.05	0.08	0.27
	(0.01)	(0.00)	(0.00)	
Father has GCSE A-Cs or equiv.	0.33	0.23	0.29	0.45
	(0.01)	(0.01)	(0.01)	
Father has A Levels or equiv.	0.19	0.17	0.18	0.38
	(0.01)	(0.01)	(0.01)	
Father has HE below degree	0.10	0.16	0.12	0.33
	(0.01)	(0.01)	(0.00)	
Father has Degree or equiv.	0.06	0.26	0.14	0.35
	(0.00)	(0.01)	(0.01)	
KS2 Average Raw Point Score	55.62	69.39	60.98	14.67
	(0.31)	(0.27)	(0.28)	
Community Tech. KS3 College	0.01	0.01	0.01	0.09
	(0.00)	(0.01)	(0.00)	
Community KS3 School	0.70	0.54	0.64	0.48
	(0.02)	(0.02)	(0.02)	
Foundation KS3 School	0.16	0.18	0.16	0.37
	(0.02)	(0.02)	(0.02)	
Independent KS3 School	0.02	0.10	0.05	0.21
	(0.00)	(0.01)	(0.01)	
Voluntary Aided KS3 School	0.09	0.14	0.11	0.31
	(0.01)	(0.02)	(0.01)	
Voluntary Controlled KS3 School	0.03	0.04	0.03	0.17
	(0.01)	(0.01)	(0.01)	
KS3 School is Grammar/Selective	0.01	0.09	0.04	0.20
	(0.00)	(0.01)	(0.00)	
KS3 School has Sixth Form	0.56	0.60	0.58	0.49
	(0.02)	(0.02)	(0.02)	
N	4488	2858	7346	

Notes: Adjusted using LSYPE Wave 7 respondent weights. Standard errors in brackets.

B Marginal effects of regression models

Marginal effects for the regression models have been estimated at sample means. All but a handful of effects are reported; those not reported due to space limitations are all indicated at the base of each table.

For most of this commentary I concentrate on the marginal effects for household incomes below £40,000. As previously noted, less than 10% of the sample are in this upper part of the distribution. Above the £40,000 threshold the estimates are generally (but not exclusively) small and insignificant. I return to this top part of the distribution later.

A £10,000 change in household equivalised income is equivalent to a family of four's income increasing by £20,000. This is, for example, equivalent to that family's position in the income distribution shifting from the 36th percentile to the 74th percentile. It is a shift of 0.7 of a standard deviation.

Considering first the attendance models, both female (and male), we find that M1 identifies a marginal effect of 14 (13) percentage points in increased university attendance for every additional £10,000 of household income below £40,000. Comparing this with the Apply models, we find that the marginal effects have barely altered compared to the university attendance models. It comes as little surprise then that our first Conditional Attend model identifies a relatively small (but significant) association between income below £40,000 and attendance of around 6 percentage points for both males and females, even with no controlling factors.

As one would expect, and previous research suggests, these associations are much reduced once additional covariates are controlled for. The base regression model takes no account of prior attainment, which acts both as an imperfect measure of underlying ability and as a function of socioeconomic characteristics on attainment up to that point. Once KS2 attainment is included in M2 the gradient falls around a half to 7 (7.6) percentage points. The relatively small effect of income on attendance, conditional on having applied, becomes even smaller, in this case the gradient drops to 2.5 (2.6) percentage points for females (males). Although smaller, this is perhaps a more robust indication of socioeconomic variation in university attendance, once a correlate of ability has been controlled for.

We observe further drops once socioeconomic characteristics are added in M3 and the marginal effects for conditional attendance become insignificant. However, also interesting given the specification is to examine the other significant associations. There are significant marginal effects for the ethnicity dummy variables, sibling effect dummy variables and a small but significant effect of month of birth. Notably, we see significant estimated effects for lone parent family status and some parental education variables. Father having a degree, in particular, shows a large and significant marginal effect comparable to £20-30,000 of additional household equivalised income.

For M4, we return to just controlling for prior attainment, this time up to Key Stage 4/GCSE. For attendance the income gradient drops to just 2 percentage points. This is in contrast to previous research where no significant effect of socioeconomic status is generally identified once educational outcomes at the age of 16 are controlled for. These results suggest that there is a small, but significant, effect of household income after the age of 16, but much of the association from earlier models was having effect through improved educational performance earlier in the school career. We can again use the application and conditional attendance models to see that this effect seems to be driven by the application decision: for conditional attendance, below £40,000, no statistically significant effect is identified and the marginal effects are 0 to two decimal places.

Very little change in the apparent effect of income is observed when socioeconomic factors are also included in M5. Nevertheless, some of the the same parental education variables once again show a smaller, but still significant marginal effect.

In M6 and M7, the marginal effect below £40,000 is roughly 4 percentage points. It seems surprising that even within a school and for individuals with otherwise very similar socioeconomic characteristics a small association between household income and university attendance is still identified.

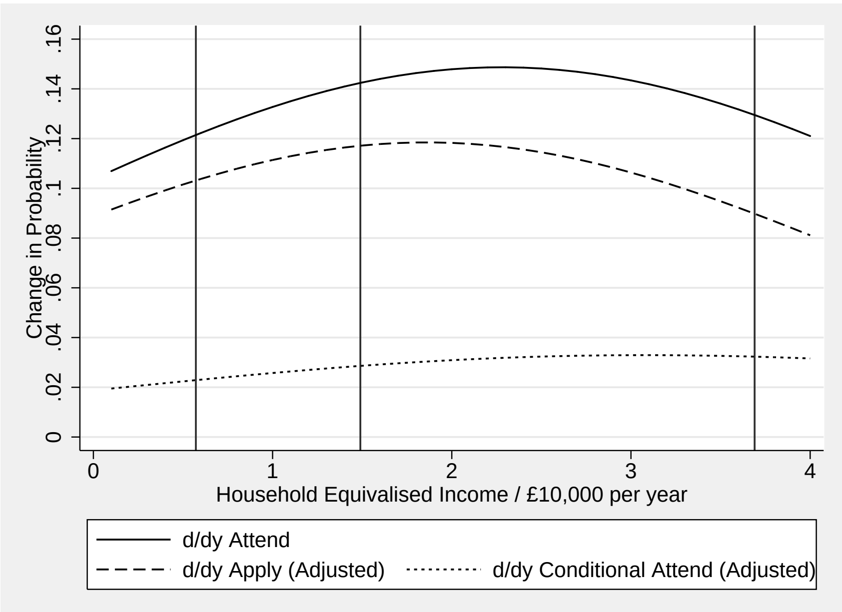
Using the conditional relationship between application and attendance, we can derive the relationship between the marginal effects of household income on application, attendance and conditional attendance. This is shown by applying the product rule to Equation 5, yielding Equation 6. Intuitively, this may be thought of as placing all marginal effects in the same scale.

$$Pr(\text{Attend}) = Pr(\text{Attend} \mid \text{Apply}) \times Pr(\text{Apply}) \tag{5}$$

$$\frac{d}{dy}Pr(\text{Attend}) = \frac{d}{dy}Pr(\text{Attend} \mid \text{Apply}) \times Pr(\text{Apply}) + \frac{d}{dy}Pr(\text{Apply}) \times Pr(\text{Attend} \mid \text{Apply}) \tag{6}$$

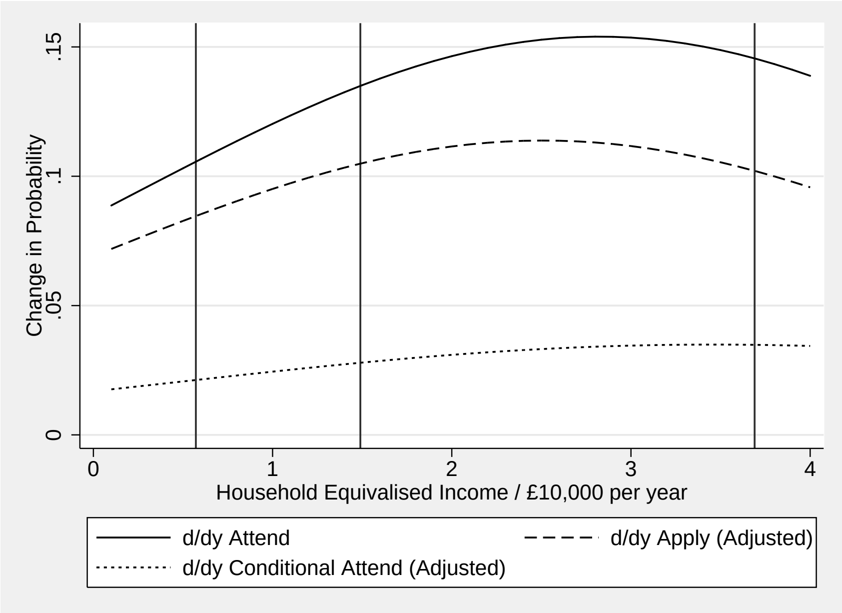
We know from our summary statistics and marginal effects estimates that the adjusted Conditional Attendance term will be smaller than the adjusted Application term. We can also confirm this by calculating these adjusted marginal effects as they vary across the income range. Figures 10 and 11 shows these for the income range £0 - 40,000. These support the message that income's association with the decision to apply to university seems to be more important than its association with whether an individual gets a place, having applied.

Figure 10: Marginal Effect of household income on university attendance for females, decomposed into application and conditional attendance effects



Notes: Estimates from Model 1 (household income only). Sample size: 4036. Vertical lines show 10th, 50th and 90th percentiles of income.

Figure 11: Marginal Effect of household income on university attendance for males, decomposed into application and conditional attendance effects



Notes: Estimates from Model 1 (household income only). Sample size: 3812. Vertical lines show 10th, 50th and 90th percentiles of income.

B.1 Russell Group

Considering the unconditional model of whether an individual attends a Russell Group university we see that for our base model there is a marginal effect of 5 percentage points for every additional £10,000 of household income. However, we know from our main analysis that the attendance gap can mask variation in how it emerges.

When we consider a model of Russell Group attendance conditional on attending any university (also see graphically in Figure 9) we see that a more pronounced gap by income remains. If an individual attends university, the base regression model demonstrates a 6 percentage point gradient for each additional £10,000 of equivalised household income.

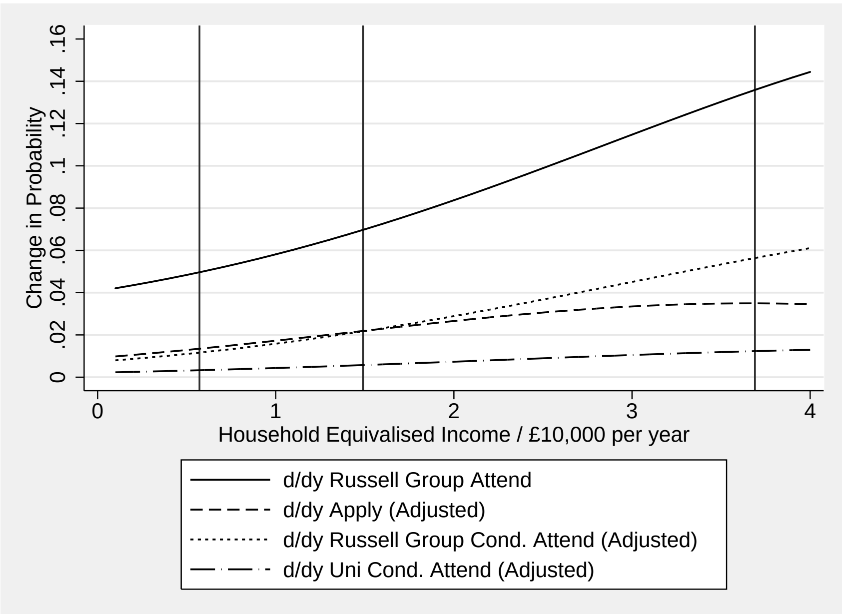
A noticeable feature of the relationship, in comparison with university conditional attendance, is that although it becomes smaller it is persistent as additional covariates are added. With Key Stage Two prior attainment added (M2) this relationship is reduced by a third to 4 percentage points. When other socioeconomic controls it falls further to 2 percentage points but does not fall further except in the fixed effects model.

Using these results along with those from the main analysis we can again decompose the overall relationship between household income and Russell Group attendance using the product rule. This relationship is shown in Equation 8. We can consider this graphically across a range of incomes, in Figure 12, and hence gain a better idea of the contribution of each to the overall observed gap.

$$Pr(\text{Russell}) = Pr(\text{Russell} \mid \text{Attend}) \times Pr(\text{Attend} \mid \text{Apply}) \times Pr(\text{Apply}) \tag{7}$$

$$\begin{aligned} \frac{d}{dy}Pr(\text{Russell}) &= \frac{d}{dy}Pr(\text{Russell} \mid \text{Attend}) \times Pr(\text{Attend} \mid \text{Apply}) \times Pr(\text{Apply}) \\ &+ Pr(\text{Russell} \mid \text{Attend}) \times \frac{d}{dy}Pr(\text{Attend} \mid \text{Apply}) \times Pr(\text{Apply}) \\ &+ Pr(\text{Russell} \mid \text{Attend}) \times Pr(\text{Attend} \mid \text{Apply}) \times \frac{d}{dy}Pr(\text{Apply}) \end{aligned} \tag{8}$$

Figure 12: Marginal Effect of household income on Russell Group attendance, decomposed into application, conditional attendance at university and conditional attendance at Russell Group effects



Notes: Estimates from Model 1 (household income only). Sample size: 7847. Vertical lines show 10th, 50th and 90th percentiles of income.

For a little over half of the income distribution the marginal effect of income on attending the Russell Group (conditional on attending university) is similar to the marginal effect on applying to university. It is certainly larger than that observed for University conditional attendance.

Table 15: Models for university attendance reporting marginal effects at means - Females

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income $\leq$ £40,000	0.142	0.070	0.050	0.019	0.027	0.039	0.032
	(0.009)***	(0.009)***	(0.010)***	(0.007)**	(0.008)***	(0.010)***	(0.010)***
Equiv. Perm. Income $>$ £40,000	0.044	0.048	0.038	0.020	0.013	0.011	0.014
	(0.023)*	(0.024)**	(0.022)*	(0.021)	(0.017)	(0.019)	(0.014)
KS2 Average Raw Point Score		-0.048	-0.010	-0.054	-0.033	0.007	-0.086
		(0.035)	(0.035)	(0.035)	(0.036)	(0.036)	(0.028)***
KS2 Average Raw Point Score Sq		0.014	0.010	0.002	0.001	0.009	0.016
		(0.003)***	(0.003)***	(0.003)	(0.003)	(0.003)***	(0.002)***
Capped GCSE Score $\leq$ 250				0.007	0.005		
				(0.003)**	(0.003)*		
Capped GCSE Score $>$ 250				0.052	0.049		
				(0.005)***	(0.005)***		
Capped GCSE Score $>$ 250 Sq				-0.000	-0.000		
				(0.000)*	(0.000)**		
Family Type Lone Parent			-0.058		0.001	-0.058	-0.048
			(0.019)***		(0.016)	(0.018)***	(0.020)**
Mother has no quals			-0.051		-0.033	-0.046	-0.040
			(0.026)**		(0.022)	(0.024)*	(0.026)
Mother has L1 or lower quals			-0.054		-0.057	-0.054	-0.052
			(0.031)*		(0.026)**	(0.030)*	(0.031)*
Mother has A Levels or equiv.			0.021		-0.000	0.020	0.002
			(0.022)		(0.018)	(0.021)	(0.026)
Mother has HE below degree			0.060		0.014	0.046	0.040
			(0.024)**		(0.020)	(0.023)**	(0.028)
Mother has Degree or equiv.			0.048		-0.003	0.040	0.026
			(0.033)		(0.026)	(0.031)	(0.033)
Father has no quals			-0.001		0.021	0.007	-0.002
			(0.025)		(0.020)	(0.024)	(0.025)
Father has L1 or lower quals			-0.008		0.004	-0.006	-0.004
			(0.033)		(0.027)	(0.032)	(0.034)
Father has A Levels or equiv.			-0.020		-0.016	-0.019	-0.018
			(0.022)		(0.019)	(0.021)	(0.026)
Father has HE below degree			0.046		0.024	0.039	0.043
			(0.026)*		(0.021)	(0.024)	(0.029)
Father has Degree or equiv.			0.127		0.069	0.110	0.099
			(0.029)***		(0.024)***	(0.028)***	(0.032)***
Community Tech. KS3 College						-0.076	
						(0.097)	
Foundation KS3 School						0.008	
						(0.023)	
Independent KS3 School						0.339	
						(0.054)***	
Voluntary Aided KS3 School						0.051	
						(0.023)**	
Voluntary Controlled KS3 School						0.056	
						(0.044)	
KS3 School is Grammar/Selective						0.152	
						(0.055)***	
KS3 School has Sixth Form						0.036	
						(0.017)**	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	163.071	18.069	171.724	25.249	14.786	21.358
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	4036	4036	4036	4036	4036	4036	4036

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 16: Models for university attendance reporting marginal effects at means - Males

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income $\leq$ £40,000	0.135	0.076	0.052	0.017	0.018	0.040	0.035
	(0.009)***	(0.008)***	(0.009)***	(0.007)**	(0.008)**	(0.010)***	(0.010)***
Equiv. Perm. Income $>$ £40,000	0.013	0.019	0.008	0.012	0.005	-0.005	-0.006
	(0.019)	(0.016)	(0.012)	(0.011)	(0.009)	(0.010)	(0.013)
KS2 Average Raw Point Score		-0.060	-0.004	-0.001	0.041	0.023	-0.104
		(0.034)*	(0.034)	(0.037)	(0.037)	(0.035)	(0.030)***
KS2 Average Raw Point Score Sq		0.015	0.010	-0.001	-0.003	0.007	0.017
		(0.003)***	(0.003)***	(0.003)	(0.003)	(0.003)**	(0.003)***
Capped GCSE Score $\leq$ 250				0.012	0.010		
				(0.005)**	(0.005)**		
Capped GCSE Score $>$ 250				0.033	0.032		
				(0.005)***	(0.005)***		
Capped GCSE Score $>$ 250 Sq				0.000	0.000		
				(0.000)	(0.000)		
Family Type Lone Parent			-0.052		-0.013	-0.052	-0.050
			(0.020)***		(0.017)	(0.020)***	(0.019)**
Mother has no quals			0.001		0.022	0.009	0.026
			(0.025)		(0.021)	(0.024)	(0.026)
Mother has L1 or lower quals			0.004		0.037	0.006	-0.001
			(0.032)		(0.027)	(0.031)	(0.030)
Mother has A Levels or equiv.			0.028		0.007	0.013	0.014
			(0.020)		(0.018)	(0.020)	(0.026)
Mother has HE below degree			0.062		0.053	0.055	0.072
			(0.024)***		(0.019)***	(0.024)**	(0.028)***
Mother has Degree or equiv.			0.063		0.020	0.054	0.046
			(0.028)**		(0.024)	(0.028)*	(0.033)
Father has no quals			0.014		0.027	0.014	0.041
			(0.025)		(0.022)	(0.024)	(0.025)
Father has L1 or lower quals			0.006		0.038	0.011	0.026
			(0.034)		(0.028)	(0.033)	(0.033)
Father has A Levels or equiv.			0.008		0.003	0.008	0.017
			(0.021)		(0.017)	(0.020)	(0.024)
Father has HE below degree			0.076		0.063	0.075	0.090
			(0.025)***		(0.020)***	(0.025)***	(0.031)***
Father has Degree or equiv.			0.168		0.092	0.158	0.192
			(0.028)***		(0.024)***	(0.027)***	(0.032)***
Community Tech. KS3 College						0.056	
						(0.055)	
Foundation KS3 School						0.006	
						(0.021)	
Independent KS3 School						0.220	
						(0.034)***	
Voluntary Aided KS3 School						0.067	
						(0.023)***	
Voluntary Controlled KS3 School						0.045	
						(0.039)	
KS3 School is Grammar/Selective						0.134	
						(0.039)***	
KS3 School has Sixth Form						0.024	
						(0.016)	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	170.682	16.132	134.546	20.836	15.934	19.167
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	3812	3812	3812	3812	3812	3812	3812

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 17: Models for university application reporting marginal effects at means - Females

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income $\leq$ £40,000	0.156	0.080	0.065	0.022	0.037	0.052	0.043
	(0.010)***	(0.010)***	(0.011)***	(0.008)***	(0.009)***	(0.011)***	(0.010)***
Equiv. Perm. Income $>$ £40,000	0.050	0.062	0.052	0.047	0.029	0.014	0.008
	(0.023)**	(0.027)**	(0.023)**	(0.023)**	(0.017)*	(0.016)	(0.011)
KS2 Average Raw Point Score		-0.120	-0.091	-0.043	-0.032	-0.090	-0.059
		(0.036)***	(0.037)**	(0.034)	(0.035)	(0.037)**	(0.033)*
KS2 Average Raw Point Score Sq		0.020	0.017	0.001	0.001	0.016	0.014
		(0.003)***	(0.003)***	(0.003)	(0.003)	(0.003)***	(0.003)***
Capped GCSE Score $\leq$ 250				0.007	0.006		
				(0.002)***	(0.002)***		
Capped GCSE Score $>$ 250				0.033	0.030		
				(0.005)***	(0.005)***		
Capped GCSE Score $>$ 250 Sq				0.001	0.001		
				(0.000)**	(0.000)**		
Family Type Lone Parent			-0.052		0.010	-0.052	-0.059
			(0.021)**		(0.018)	(0.020)***	(0.022)***
Mother has no quals			-0.047		-0.030	-0.042	-0.025
			(0.027)*		(0.022)	(0.026)	(0.028)
Mother has L1 or lower quals			-0.062		-0.063	-0.062	-0.065
			(0.030)**		(0.025)**	(0.029)**	(0.033)*
Mother has A Levels or equiv.			0.003		-0.020	0.000	-0.008
			(0.026)		(0.022)	(0.025)	(0.027)
Mother has HE below degree			0.076		0.032	0.062	0.061
			(0.027)***		(0.023)	(0.025)**	(0.028)**
Mother has Degree or equiv.			0.062		0.005	0.059	0.038
			(0.037)*		(0.031)	(0.036)	(0.030)
Father has no quals			-0.010		0.011	-0.002	-0.022
			(0.027)		(0.022)	(0.026)	(0.028)
Father has L1 or lower quals			-0.018		-0.002	-0.014	-0.020
			(0.034)		(0.027)	(0.033)	(0.036)
Father has A Levels or equiv.			-0.001		0.004	0.001	-0.022
			(0.026)		(0.023)	(0.025)	(0.027)
Father has HE below degree			0.042		0.015	0.035	0.048
			(0.029)		(0.025)	(0.028)	(0.029)*
Father has Degree or equiv.			0.115		0.062	0.094	0.046
			(0.033)***		(0.029)**	(0.033)***	(0.030)
Community Tech. KS3 College						-0.114	
						(0.147)	
Foundation KS3 School						0.026	
						(0.025)	
Independent KS3 School						0.427	
						(0.063)***	
Voluntary Aided KS3 School						0.063	
						(0.026)**	
Voluntary Controlled KS3 School						0.073	
						(0.045)	
KS3 School is Grammar/Selective						0.085	
						(0.084)	
KS3 School has Sixth Form						0.048	
						(0.017)***	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	137.469	14.904	133.523	20.363	14.379	21.249
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	4036	4036	4036	4036	4036	4036	4036

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 18: Models for university application reporting marginal effects at means - Males

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income $\leq$ £40,000	0.147	0.086	0.055	0.019	0.016	0.039	0.034
	(0.010)***	(0.009)***	(0.010)***	(0.007)**	(0.009)*	(0.010)***	(0.010)***
Equiv. Perm. Income > £40,000	-0.003	0.005	-0.011	-0.004	-0.015	-0.028	-0.023
	(0.019)	(0.017)	(0.012)	(0.012)	(0.011)	(0.012)**	(0.013)*
KS2 Average Raw Point Score		-0.119	-0.060	0.008	0.041	-0.038	-0.035
		(0.038)***	(0.038)	(0.036)	(0.036)	(0.037)	(0.033)
KS2 Average Raw Point Score Sq		0.019	0.014	-0.003	-0.004	0.011	0.011
		(0.003)***	(0.003)***	(0.003)	(0.003)	(0.003)***	(0.003)***
Capped GCSE Score $\leq$ 250				0.009	0.007		
				(0.002)***	(0.002)***		
Capped GCSE Score > 250				0.022	0.022		
				(0.005)***	(0.005)***		
Capped GCSE Score > 250 Sq				0.001	0.001		
				(0.000)***	(0.000)***		
Family Type Lone Parent			-0.073		-0.022	-0.072	-0.078
			(0.022)***		(0.018)	(0.021)***	(0.022)***
Mother has no quals			0.008		0.033	0.015	0.006
			(0.028)		(0.023)	(0.026)	(0.028)
Mother has L1 or lower quals			0.009		0.037	0.011	-0.010
			(0.031)		(0.027)	(0.030)	(0.034)
Mother has A Levels or equiv.			0.045		0.018	0.026	0.018
			(0.024)*		(0.021)	(0.023)	(0.027)
Mother has HE below degree			0.070		0.057	0.056	0.075
			(0.026)***		(0.022)***	(0.026)**	(0.029)***
Mother has Degree or equiv.			0.096		0.043	0.082	0.069
			(0.031)***		(0.027)	(0.030)***	(0.031)**
Father has no quals			-0.021		-0.005	-0.021	-0.002
			(0.026)		(0.022)	(0.025)	(0.028)
Father has L1 or lower quals			-0.020		0.011	-0.013	-0.005
			(0.034)		(0.027)	(0.032)	(0.037)
Father has A Levels or equiv.			-0.015		-0.019	-0.016	0.014
			(0.023)		(0.019)	(0.021)	(0.026)
Father has HE below degree			0.071		0.058	0.069	0.090
			(0.028)**		(0.024)**	(0.027)**	(0.032)***
Father has Degree or equiv.			0.188		0.105	0.171	0.177
			(0.031)***		(0.027)***	(0.030)***	(0.030)***
Community Tech. KS3 College						0.072	
						(0.086)	
Foundation KS3 School						-0.008	
						(0.025)	
Independent KS3 School						0.327	
						(0.057)***	
Voluntary Aided KS3 School						0.078	
						(0.027)***	
Voluntary Controlled KS3 School						-0.004	
						(0.043)	
KS3 School is Grammar/Selective						0.127	
						(0.037)***	
KS3 School has Sixth Form						0.056	
						(0.017)***	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	138.130	14.825	126.328	19.562	15.715	18.778
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	3812	3812	3812	3812	3812	3812	3812

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 19: Models for university attendance, conditional on application, reporting marginal effects at means - Females

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income < £40,000	0.056	0.025	0.014	0.005	0.007	0.010	0.001
	(0.011)***	(0.010)**	(0.012)	(0.009)	(0.011)	(0.011)	(0.011)
Equiv. Perm. Income > £40,000	0.026	0.028	0.021	0.015	0.010	0.008	0.010
	(0.024)	(0.023)	(0.022)	(0.021)	(0.020)	(0.021)	(0.012)
KS2 Average Raw Point Score		0.034	0.049	-0.037	-0.035	0.066	0.096
		(0.047)	(0.046)	(0.052)	(0.052)	(0.046)	(0.058)
KS2 Average Raw Point Score Sq		0.004	0.002	0.002	0.002	0.000	-0.002
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Capped GCSE Score < 250				0.007	0.007		
				(0.006)	(0.006)		
Capped GCSE Score > 250				0.038	0.038		
				(0.006)***	(0.006)***		
Capped GCSE Score > 250 Sq				-0.000	-0.001		
				(0.000)*	(0.000)*		
Family Type Lone Parent			-0.038		-0.003	-0.037	-0.033
			(0.025)		(0.023)	(0.024)	(0.027)
Mother has no quals			-0.041		-0.034	-0.039	-0.033
			(0.033)		(0.031)	(0.033)	(0.041)
Mother has L1 or lower quals			-0.015		-0.026	-0.017	-0.009
			(0.043)		(0.040)	(0.043)	(0.053)
Mother has A Levels or equiv.			0.036		0.021	0.037	0.018
			(0.029)		(0.026)	(0.029)	(0.032)
Mother has HE below degree			0.014		-0.008	0.008	-0.003
			(0.029)		(0.027)	(0.029)	(0.032)
Mother has Degree or equiv.			0.015		-0.007	0.008	-0.011
			(0.037)		(0.033)	(0.035)	(0.036)
Father has no quals			-0.005		-0.001	-0.000	0.003
			(0.032)		(0.030)	(0.032)	(0.040)
Father has L1 or lower quals			-0.007		-0.014	-0.007	0.018
			(0.042)		(0.039)	(0.040)	(0.055)
Father has A Levels or equiv.			-0.049		-0.043	-0.047	-0.019
			(0.030)		(0.029)	(0.030)	(0.033)
Father has HE below degree			0.020		0.022	0.014	-0.013
			(0.034)		(0.031)	(0.032)	(0.036)
Father has Degree or equiv.			0.078		0.048	0.071	0.072
			(0.035)**		(0.033)	(0.035)**	(0.033)**
Community Tech. KS3 College						-0.006	
						(0.113)	
Foundation KS3 School						-0.022	
						(0.025)	
Independent KS3 School						0.158	
						(0.053)***	
Voluntary Aided KS3 School						0.014	
						(0.025)	
Voluntary Controlled KS3 School						0.020	
						(0.049)	
KS3 School is Grammar/Selective						0.160	
						(0.051)***	
KS3 School has Sixth Form						0.002	
						(0.020)	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	34.079	4.341	47.147	7.640	3.736	3.881
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	2638	2638	2638	2638	2638	2638	2638

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 20: Models for university attendance, conditional on application, reporting marginal effects at means - Males

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income < £40,000	0.066	0.026	0.022	0.004	0.010	0.021	0.017
	(0.014)***	(0.012)**	(0.014)	(0.012)	(0.013)	(0.014)	(0.013)
Equiv. Perm. Income > £40,000	0.094	0.092	0.071	0.067	0.051	0.060	0.011
	(0.040)**	(0.038)**	(0.035)**	(0.032)**	(0.029)*	(0.035)*	(0.013)
KS2 Average Raw Point Score		0.048	0.085	0.005	0.049	0.114	0.222
		(0.059)	(0.058)	(0.062)	(0.062)	(0.058)**	(0.068)***
KS2 Average Raw Point Score Sq		0.005	0.002	0.000	-0.003	-0.001	-0.010
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)*
Capped GCSE Score < 250				0.018	0.016		
				(0.011)*	(0.010)		
Capped GCSE Score > 250				0.024	0.025		
				(0.008)***	(0.008)***		
Capped GCSE Score > 250 Sq				0.000	0.000		
				(0.000)	(0.000)		
Family Type Lone Parent			-0.017		0.002	-0.016	-0.038
			(0.031)		(0.029)	(0.031)	(0.037)
Mother has no quals			-0.003		-0.004	-0.000	-0.009
			(0.038)		(0.034)	(0.038)	(0.047)
Mother has L1 or lower quals			0.005		0.045	0.006	0.029
			(0.050)		(0.046)	(0.049)	(0.059)
Mother has A Levels or equiv.			0.003		0.005	-0.004	0.023
			(0.031)		(0.029)	(0.031)	(0.038)
Mother has HE below degree			0.044		0.035	0.041	0.060
			(0.036)		(0.032)	(0.036)	(0.038)
Mother has Degree or equiv.			0.018		0.005	0.014	0.021
			(0.038)		(0.037)	(0.037)	(0.039)
Father has no quals			0.064		0.065	0.063	0.088
			(0.038)*		(0.035)*	(0.037)*	(0.045)*
Father has L1 or lower quals			0.064		0.065	0.070	0.045
			(0.054)		(0.049)	(0.053)	(0.064)
Father has A Levels or equiv.			0.040		0.018	0.037	0.018
			(0.033)		(0.030)	(0.032)	(0.039)
Father has HE below degree			0.069		0.070	0.069	0.057
			(0.037)*		(0.033)**	(0.037)*	(0.042)
Father has Degree or equiv.			0.115		0.070	0.114	0.089
			(0.039)***		(0.037)*	(0.038)***	(0.039)**
Community Tech. KS3 College						0.006	
						(0.082)	
Foundation KS3 School						0.014	
						(0.029)	
Independent KS3 School						0.049	
						(0.059)	
Voluntary Aided KS3 School						0.038	
						(0.029)	
Voluntary Controlled KS3 School						0.103	
						(0.060)*	
KS3 School is Grammar/Selective						0.129	
						(0.053)**	
KS3 School has Sixth Form						-0.025	
						(0.023)	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	39.852	3.561	37.960	5.962	3.566	3.607
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	2207	2207	2207	2207	2207	2207	2207

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Notes:*Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 21: Models for Russell Group university attendance reporting marginal effects at means

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income < £40,000	0.047	0.028	0.017	0.012	0.009	0.013	0.014
	(0.003)***	(0.002)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.005)***
Equiv. Perm. Income > £40,000	0.004	0.004	0.002	-0.003	-0.003	-0.003	-0.003
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.010)
KS2 Average Raw Point Score		-0.075	-0.061	-0.041	-0.029	-0.046	-0.174
		(0.017)***	(0.017)***	(0.022)*	(0.021)	(0.019)**	(0.014)***
KS2 Average Raw Point Score Sq		0.011	0.009	0.003	0.002	0.008	0.019
		(0.001)***	(0.001)***	(0.002)*	(0.002)	(0.001)***	(0.001)***
Capped GCSE Score < 250				0.003	0.003		
				(0.002)**	(0.001)*		
Capped GCSE Score > 250				0.008	0.008		
				(0.003)**	(0.003)**		
Capped GCSE Score > 250 Sq				0.000	0.000		
				(0.000)***	(0.000)***		
Family Type Lone Parent			-0.005		0.015	-0.005	-0.000
			(0.009)		(0.009)*	(0.009)	(0.008)
Mother has no quals			0.004		0.008	0.009	0.011
			(0.012)		(0.011)	(0.011)	(0.009)
Mother has L1 or lower quals			-0.010		-0.007	-0.006	0.000
			(0.016)		(0.015)	(0.016)	(0.009)
Mother has A Levels or equiv.			0.015		0.009	0.013	0.001
			(0.010)		(0.010)	(0.010)	(0.011)
Mother has HE below degree			0.047		0.039	0.044	0.057
			(0.010)***		(0.009)***	(0.010)***	(0.013)***
Mother has Degree or equiv.			0.042		0.024	0.041	0.050
			(0.010)***		(0.009)***	(0.010)***	(0.019)***
Father has no quals			0.025		0.032	0.028	0.034
			(0.012)**		(0.012)***	(0.012)**	(0.009)***
Father has L1 or lower quals			0.013		0.024	0.016	0.026
			(0.016)		(0.015)	(0.016)	(0.011)**
Father has A Levels or equiv.			0.021		0.018	0.021	0.017
			(0.010)**		(0.009)**	(0.010)**	(0.010)*
Father has HE below degree			0.018		0.012	0.019	0.006
			(0.011)		(0.010)	(0.011)*	(0.013)
Father has Degree or equiv.			0.060		0.038	0.058	0.091
			(0.010)***		(0.009)***	(0.010)***	(0.017)***
Community Tech. KS3 College						0.014	
						(0.010)	
Foundation KS3 School						0.001	
						(0.010)	
Independent KS3 School						0.083	
						(0.014)***	
Voluntary Aided KS3 School						0.022	
						(0.011)*	
Voluntary Controlled KS3 School						0.009	
						(0.015)	
KS3 School is Grammar/Selective						0.042	
						(0.013)***	
KS3 School has Sixth Form						0.008	
						(0.008)	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	185.898	19.041	185.030	30.102	17.675	11.597
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	7847	7847	7847	7847	7847	7847	7847

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

Table 22: Models for Russell Group attendance, conditional on attending university, reporting marginal effects at means

	M1-Probit	M2-Probit	M3-Probit	M4-Probit	M5-Probit	M6-Probit	M7-LP (with FE)
Equiv. Perm. Income < £40,000	0.063	0.042	0.025	0.024	0.018	0.019	0.012
	(0.006)***	(0.007)***	(0.009)***	(0.007)***	(0.009)**	(0.009)**	(0.010)
Equiv. Perm. Income > £40,000	0.005	0.002	-0.000	-0.009	-0.008	-0.007	-0.007
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.013)
KS2 Average Raw Point Score		-0.249	-0.216	-0.094	-0.067	-0.189	-0.303
		(0.063)***	(0.061)***	(0.070)	(0.066)	(0.063)***	(0.056)***
KS2 Average Raw Point Score Sq		0.027	0.024	0.008	0.005	0.022	0.031
		(0.005)***	(0.005)***	(0.005)	(0.005)	(0.005)***	(0.005)***
Capped GCSE Score < 250				-0.000	-0.001		
				(0.007)	(0.007)		
Capped GCSE Score > 250				-0.005	-0.005		
				(0.009)	(0.009)		
Capped GCSE Score > 250 Sq				0.001	0.001		
				(0.000)***	(0.000)***		
Family Type Lone Parent			0.010		0.039	0.011	0.003
			(0.024)		(0.022)*	(0.024)	(0.024)
Mother has no quals			0.038		0.024	0.045	0.058
			(0.031)		(0.029)	(0.030)	(0.032)*
Mother has L1 or lower quals			-0.011		-0.006	-0.004	-0.001
			(0.043)		(0.038)	(0.042)	(0.036)
Mother has A Levels or equiv.			0.023		0.017	0.023	0.011
			(0.025)		(0.024)	(0.025)	(0.027)
Mother has HE below degree			0.098		0.089	0.095	0.112
			(0.024)***		(0.023)***	(0.024)***	(0.028)***
Mother has Degree or equiv.			0.090		0.058	0.091	0.083
			(0.026)***		(0.023)**	(0.026)***	(0.032)***
Father has no quals			0.049		0.065	0.054	0.040
			(0.030)		(0.029)**	(0.029)*	(0.030)
Father has L1 or lower quals			0.059		0.054	0.057	0.061
			(0.042)		(0.039)	(0.042)	(0.040)
Father has A Levels or equiv.			0.052		0.042	0.050	0.037
			(0.026)**		(0.025)*	(0.025)**	(0.027)
Father has HE below degree			0.020		0.015	0.021	0.004
			(0.027)		(0.026)	(0.026)	(0.030)
Father has Degree or equiv.			0.093		0.067	0.091	0.086
			(0.025)***		(0.024)***	(0.025)***	(0.030)***
Community Tech. KS3 College						0.045	
						(0.037)	
Foundation KS3 School						0.003	
						(0.023)	
Independent KS3 School						0.124	
						(0.031)***	
Voluntary Aided KS3 School						0.031	
						(0.027)	
Voluntary Controlled KS3 School						0.013	
						(0.033)	
KS3 School is Grammar/Selective						0.065	
						(0.027)**	
KS3 School has Sixth Form						0.010	
						(0.019)	
Region	No	No	Yes	No	Yes	Yes	No
Ethnicity	No	No	Yes	No	Yes	Yes	Yes
Month of Birth	No	No	Yes	No	Yes	Yes	Yes
Sibling effects	No	No	Yes	No	Yes	Yes	Yes
F Test	.	77.625	8.209	97.383	15.624	7.834	7.023
Prob > F	.	0.000	0.000	0.000	0.000	0.000	0.000
Sub-sample size	3826	3826	3826	3826	3826	3826	3826

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Marginal effects estimated at sample means. Weighted using Wave 7 Participant Weights, which attempt to adjust for oversampling and attrition. Standard errors are adjusted for school level clustering and stratification by deprivation. Income variables are divided by 10000, hence the coefficient estimates represent the expected change in probability for an additional £10000 of equivalised 'permanent' income. Prior attainment variables are divided by 10, hence the coefficient estimates represent the expected change in probability for an additional 10 points. Base category for parental education is achieving GCSEs or equivalents. Base category for family type is married or cohabiting couple. Base category for KS3 School Type is Community School. Sample: Wave 7 Participants with valid responses for all variables used in models. Marginal effects for discrete variables is the change from base category.

C Continuous regressors' functional form

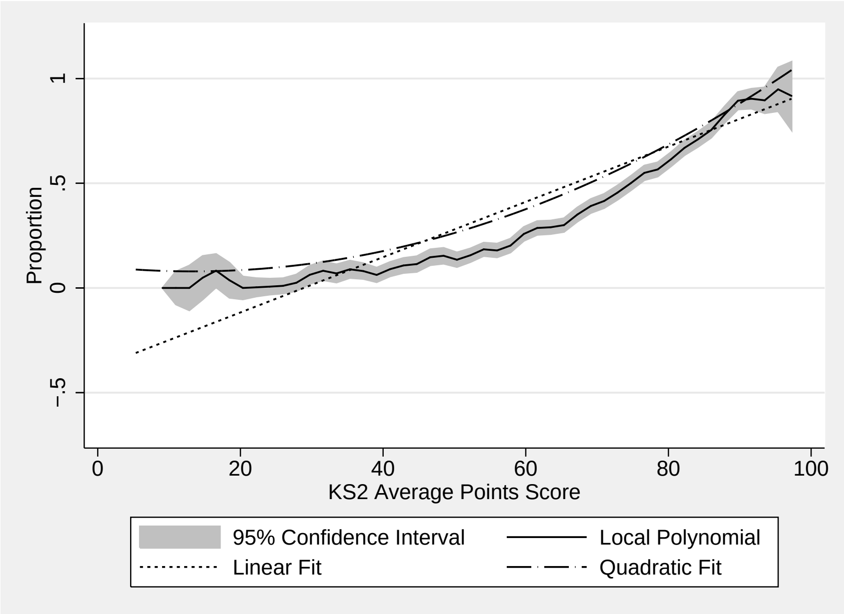
Since probit models require an assumption of normally distributed errors, specifying the appropriate functional form of continuous regressors is particularly important to reducing bias. Other than household income (discussed in Section 5) KS2 Raw Average Point score and GCSE Capped Point score are my only two such continuous regressors. In order to consider appropriate functional form within a binary choice model I follow the following procedure. I undertake local polynomial smoothing, using an Epanechnikov kernel and Silverman's optimal bandwidth. On to these were placed best linear fit and best quadratic fit lines. These are shown in Figures 13, 14 and 15. This allows visual inspection of the respective fits of using quadratic or linear terms.

For KS2 Average Point score, I judge that use of a quadratic term provides a better fit.

For GCSE attainment, neither provides a particularly good fit. It would appear that such is the importance, either directly or indirectly, of GCSE results on whether individuals apply to and attend university a capped point score of anything less than 250 results in negligible probability of these outcomes. Once past this hurdle the proportions applying and attending increase with a resulting functional form somewhat like a normal cumulative density function. I model this as a combination of piecewise linear and quadratic, with a simple linear term up to 250 points and then allow for a new quadratic term from this to the top.

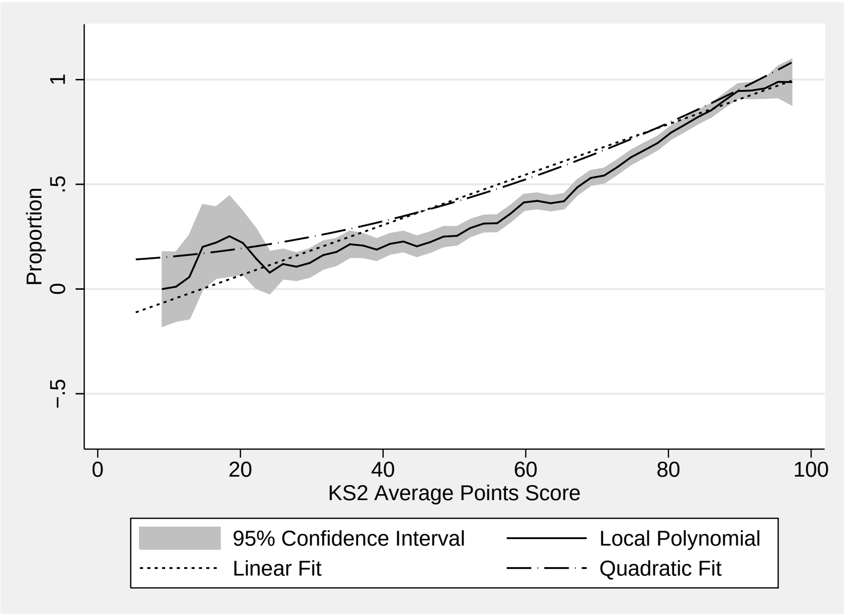
This analysis does not, of course, provide a definitive functional form for these relationships in models including other confounding factors. As such, while I use these outcomes as indicative, and hence use in initial models, I recheck relative fits and revise if appropriate.

Figure 13: Linear and Quadratic Fits of Key Stage 2 Attainment with University Attendance



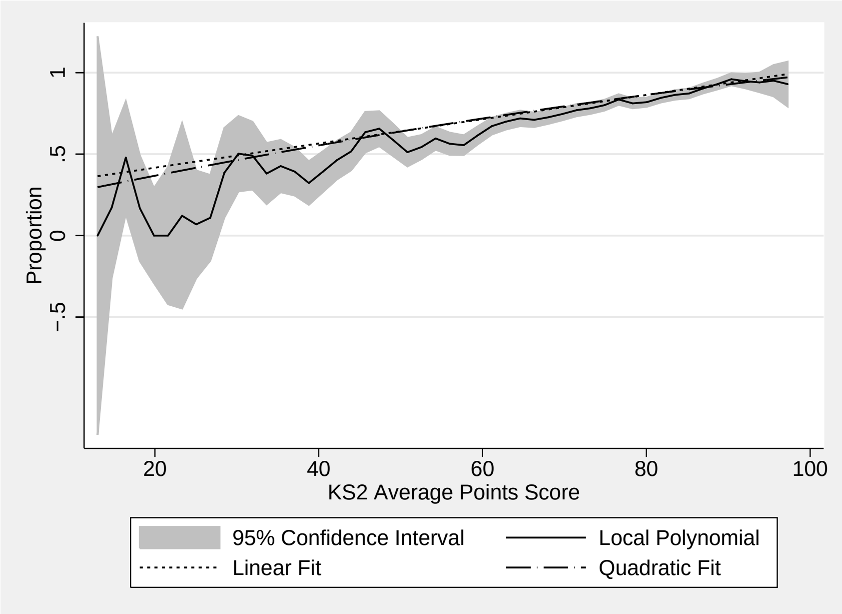
Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 1.1824005. Sample size: 8166

Figure 14: Linear and Quadratic Fits of Key Stage 2 Attainment with University Application



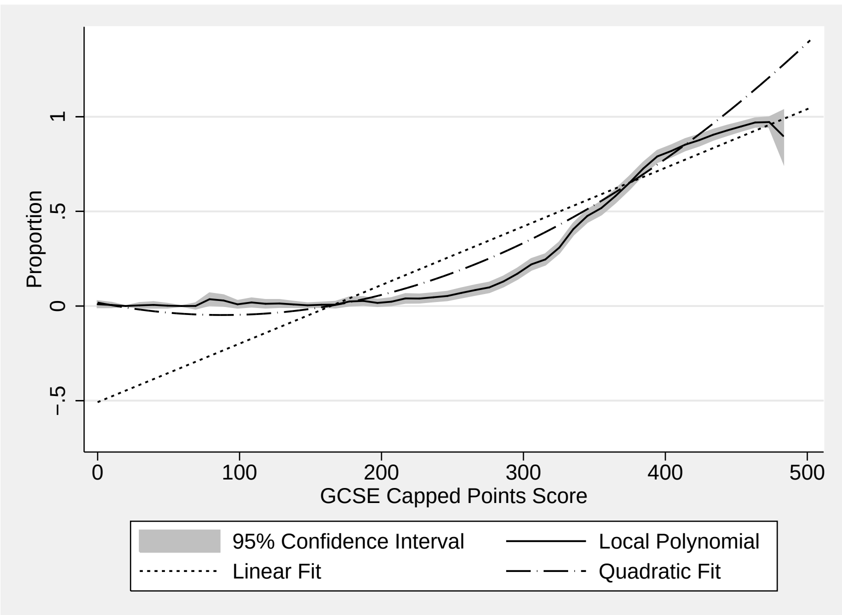
Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 1.1824005. Sample size: 8166

Figure 15: Linear and Quadratic Fits of Key Stage 2 Attainment with University Attendance, conditional on Application



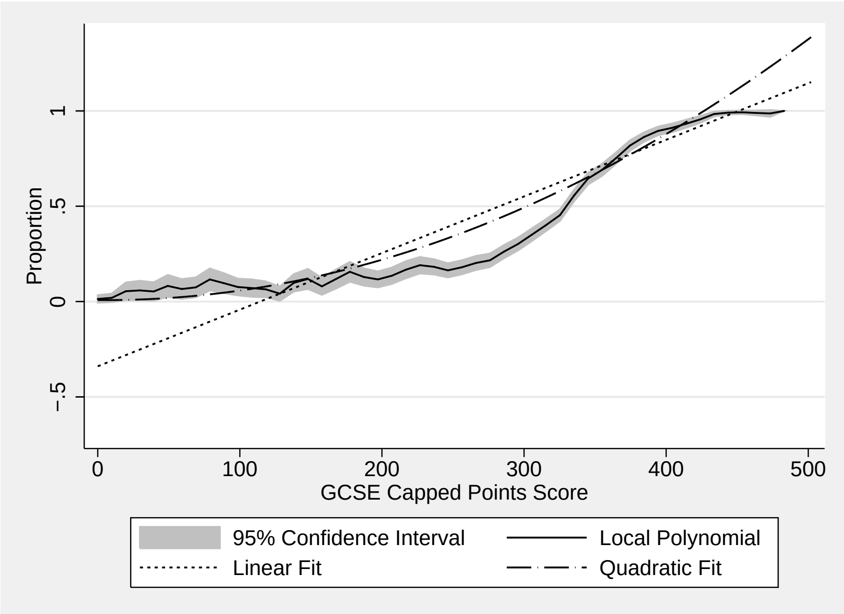
Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 1.0013723. Sample size: 5033

Figure 16: Linear and Quadratic Fits of GCSE Attainment with University Attendance



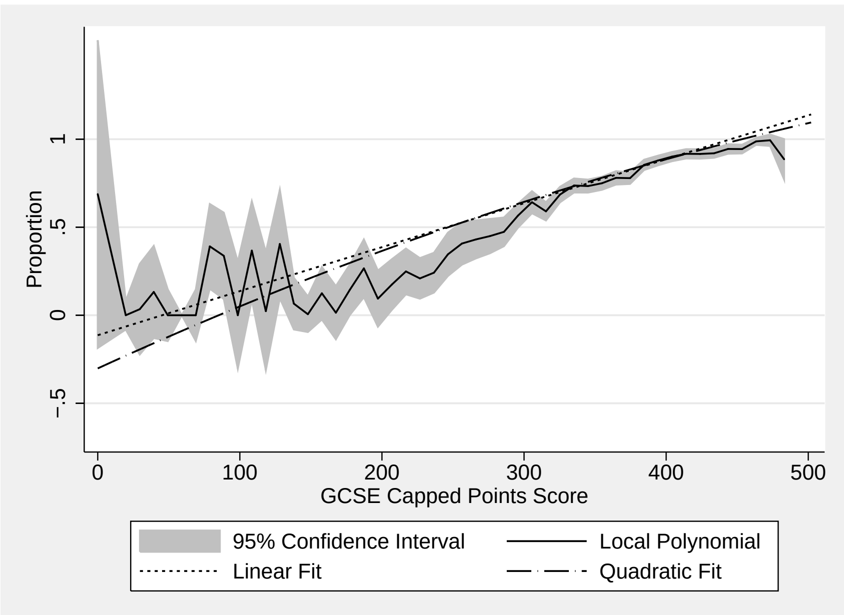
Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 13.157616. Sample size: 8624

Figure 17: Linear and Quadratic Fits of GCSE Attainment with University Application



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 13.157616. Sample size: 8624

Figure 18: Linear and Quadratic Fits of GCSE Attainment with University Attendance, conditional on Application



Notes: Weighted using LSYPE Wave 7 Participant Weights. Local polynomial smoothing using Epanechnikov kernel and Silverman's optimal bandwidth of 9.4667053. Sample size: 5307