ACCESS, CHOICE AND PARTICIPATION IN HIGHER EDUCATION

Stephen Gibbons*, Anna Vignoles**

August 2008

Author for correspondence:
Stephen Gibbons,
Department of Geography and Environment,
London School of Economics,
Houghton Street, WC2A 2AE, London, UK.
Tel.: +44 [0]20 79556245
Email: s.gibbons@lse.ac.uk

*Department of Geography and Environment, London School of Economics.
**Institute of Education

Acknowledgements: We thank Alejandra Castrodad-Rodriguez for her hard work computing the home-HE rail network distances, Anushri Bansal and Claire Crawford, for further assistance in setting up the data. The project was financed by the ESRC under the TLRP programme. The work was also supported by the Centre for the Economics of Education, London School of Economics. Data was made available by the Higher Education Statistics Agency and the Department of Children Schools and Families.
Abstract

Geographical distance between parental home and college poses a potential barrier to higher education (HE) entry, and could be a deciding factor when choosing between institutions. In particular, those on low incomes or with strong family commitments may be constrained in their education choices because they cannot afford to leave home, or have personal or cultural reasons to remain close to their family. Previous work has proposed that these factors contribute to patterns of HE participation amongst ethnic minorities and those from low income families. This paper provides quantitative evidence on these issues using administrative data on a cohort of university entrants in England merged with individual and school level information. Our findings are that geographical accessibility of HE institutions has little or no impact on the decision to participate in HE. However, geographical accessibility has a strong influence on institutional choice. The elasticity of the probability of attendance with respect to home-institution distance is around -1, and there are small, but behaviourally important differences between ethnic, occupation and income groups.

Keywords: Higher education, choice, migration

JEL Classifications: I2 R
1. Introduction

Commuting or re-location location costs can impose high barriers to university entrance, particularly for lower-income students, or students for whom there are cultural incentives to remain in or close to the parental home. Although, anecdotally, this has an important influence on higher education (college)¹ choices – and hence on earnings and life chances – there is relatively little good empirical evidence for the UK or internationally.

Economic theory and evidence on migration and commuting suggests that greater distance increases direct, informational and psychic costs of travel and learning, and makes it likely that people choose destinations close to home. In the case of university enrolment in England, this relationship is not self-evident, because a long move away from home has traditionally been seen as part of the university experience, and part of the transition to adulthood. At least this was the case for the predominantly higher income students who historically enrolled in university. As participation by previously under-represented groups, such as lower income students, has increased, so expectations about the nature of the university experience may have changed. Distance from home could be an important factor affecting institution choice amongst these ‘non-traditional’ students, and increasingly important for all groups given the disincentive to leave home implied by rising housing costs and recent increases in university tuition costs in England². Moreover, when migration or commuting is for a specific activity, like Higher Education (HE) participation, it is plausible that a lack of local institutions could make any form of participation less likely, especially amongst those who already face high costs.

¹ Unless otherwise stated, this paper uses the terms university, college and higher education participation to refer specifically to enrolment in 3 year Bachelor degrees.

² Historically, students have not paid tuition fees to attend university or college in England. Fees, payable in advance, were introduced in 1998 but were initially kept very low, at around £1,000 per annum with poorer students exempt. Students now pay around £3,000 per annum via an income contingent loan. This is set to change as universities are now empowered to charge higher fees and to vary fees by subject.
In labour economics, a familiar and analogous idea is that distance to jobs deters participation or search in the labour market (see Gobillon, et al. 2005 for a recent survey). Some limited international research has supported this idea that people living further away from universities are less likely to choose to enrol in university, more likely to attend local colleges, and that this distance ‘discount’ is higher for those from more disadvantaged backgrounds (e.g. Frenette 2004, 2006; Spiess and Wrohlich 2008). However, there is very little solid evidence for the UK or European context, and almost no international evidence on the effects of distance on institution choice specifically.

If distance matters for HE choices, then the local availability of institutions could have important consequences. Firstly, inevitable geographical disparities in university accessibility could imply that pupils who live close to universities are more likely to enrol. As well as leading to individual inequalities in human capital accumulation, this process could lead in turn to geographical disparities (city, region etc.) if students tend also to be quite immobile when they leave university. Secondly, if distance matters more for some groups of pupils (e.g. low income, ethnic minority) than it does for others, then the distribution of characteristics amongst those enrolling in HE will not reflect the distribution of characteristics amongst suitably qualified school-leavers: equilibrium take up of HE amongst school-leavers for whom distance imposes low costs (e.g. the rich) will be greater than amongst school leavers with equivalent credentials who face high distance costs (e.g. the poor). Distance from an HE institution may also affect pupils’ effort and achievement in school, if they see themselves as not being able to access a university, again reinforcing inequalities. This story, coupled with rising real re-location costs could partly explain why educational and income mobility has shown signs of decrease in England over the past decades (Blanden and Machin 2007), even though overall HE participation has expanded dramatically.

3. Interestingly, the idea that proximity encourages college attendance has also been used as a strategy in empirical work on the labour market returns to education (Card 1995).
Even if distance has no effect on participation, it could affect choice of institution, which may mean that a school-leaver from a disadvantaged background is less likely to enrol in a top-quality university than a school-leaver with identical credentials from a wealthier background, if, for example, top-ranked universities are on average further away from family homes. This will have long term consequences for wage inequality, given recent evidence of the significant wage premium earned by graduates from elite universities (Hussain et al. 2008). Therefore, evidence of important interactions between family background and distance discounts supports a case for policy to reduce these costs for low-income students, for example, by promoting geographical dispersion of top-rank university establishments, subsiding relocation costs, or extending the role of distance learning. Note that this is an efficiency issue, as well as an equity one, because students deterred by distance from university or high quality university attendance may have high economic returns to a university education, but be unable to offset the current costs of re-location against future labour market earnings.

The aim of our research is to examine empirically the role of distance in higher education participation and institution choice, and to assess whether home to institution distance matters more for some groups of students than others. We focus on three aspects of student background that dominate the ‘widening participation’ debate: gender, ethnicity and income/occupational group. The rest of this paper is organised as follows. In Section 2 we discuss the UK and international literature that has considered geographical accessibility of HE and its consequences. In Section 3 we discuss the methods we will use to analyse the issues and the administrative data on the English students on which we will apply them. Our results are split into two parts. In Section 4 we present an empirical analysis of the role of HE accessibility on the decision to participate in HE, whilst in Section 5 we focus on the role of distance in choice of institution. Section 6 concludes.

2. Literature

There is a large and growing literature on widening participation in HE and it is largely focused on:

- barriers to entry and participation experienced by non traditional students and
the extent to which the current policy framework facilitates or deters participation in HE by historically under represented groups (e.g. Jones and Thomas, 2005).

A number of important barriers to participation in HE have been widely cited in the literature, although there is no consensus on which factor is most important. Geographical distance to a higher education institution is one such potential factor (Dearing, 1997 and Gorard and Smith, 2006). The distribution of universities in the UK is not spatially even (see Tight, 2007 and we provide more evidence on this in our empirical work below). Other research has also suggested that non-traditional students, particularly first generation entrants, mature and ethnic minority applicants cite the location of a higher education institution (HEI) and its distance from their home as important in determining their decision to participate (Thomas and Quinn, 2007; Christie et al 2005; Connor et al. 1999). However, the quantitative evidence base on the relationship between the student’s geographical location and their participation or achievement in HE is limited. Sa et al. (2004) investigated the link between proximity to a higher education institution and HE participation in the Netherlands. They found prior attainment and other personal characteristics to be more important than proximity in determining HE participation. This is perhaps unsurprising given the density of higher education institutions in the Netherlands and the fact that over 90% of those graduating from secondary school


5 A number of studies have looked at the relationship between proximity to an HEI and student HE outcomes (drop out and degree classification or grade achieved). For example, a study for the UK by Johnes (1990) used data on the 1979 entry cohort to Lancaster University to examine the determinants of non-completion. She identified students’ academic prior achievement, their preparation for HE and their parental social class as being particularly important in determining drop out. However, she also found that students whose homes were close to the university (i.e. generally poorer students who lived at home during their university studies) were more likely to drop out. This may of course not indicate that living at home is detrimental to students’ studies but rather that the types of student who live at home struggle more with their higher education. This finding in particular is consistent with evidence from Woodward and Bradshaw (1989) and Johnes and McNabb (2004).
with a diploma (and therefore qualified to enter HE) go on to higher education. Frenette (2004, 2006) uses similar methods to analyze higher education participation and choice (college versus university) in Canada, where HE is rather sparse in some rural and remote regions, and finds that increased distance between home and university is associated with lower participation in HE and a greater tendency to attend local colleges rather than university. Speiss and Wrohlich (2008) have similar findings for Germany. Frenette (2006) also found that the deterrent effects of distance are stronger for lower income families, although the models used have no controls for students’ academic background.

Faggian et al. (2006; 2007) modelled the decision to migrate for university and the subsequent decision to migrate for employment. Although the focus of this work is not the HE participation decision per se, it does shed light on the sequential migration patterns of UK university students, highlighting the gender differences (Faggian et al. 2007) and ethnic differences (Faggian et al. 2006) in these patterns. Faggian et al. 2007 in particular uses various modelling approaches, including a hybrid conditional logit model to allow for both the characteristics of the individual and the nature of the sequential choices available to the individual. Controlling for some (but by no means all) aspects of individuals’ human capital and local labour market conditions, the paper confirms the previous findings in the literature that firstly those with more human capital (measured after graduation) are more likely to subsequently migrate, and secondly that those who have already migrated (by moving away to university) are more likely to be migrate again (e.g. by migrating to employment). Interestingly, the paper suggests that U.K. female graduates migrate more post-graduation than do male graduates (although females are less likely to migrate to attend HE). Faggian et al. 2006 also finds differences in migratory patterns across different ethnic groups of graduates. However, this research, by necessity, relied on relatively sparse data that does not contain full information on prior achievement of students.

In our paper, which focuses on the HE participation and institution choice specifically, we are able to go much further than previous work in specifying rich models which control for students’ human capital prior to the HE decision, and we incorporate information on the proximity not just of the
university chosen by the individual but also the proximity of other HEIs to the student’s original home location. We also estimate models of student choice between all the major individual HE institutions in England. The next section describes our data and estimation strategy.

3. Methods and data

The goal of this research is to estimate the relationship between home-university distance and students’ higher education decisions. The methodological framework measures the sensitivity of individuals’ decisions to the distance between their parental home and higher education institutions. More specifically, we consider a) the statistical association between the proximity of HE institutions to a student’s home and their decision to participate in HE at Bachelors degree level; and, b) the statistical association between distance from home to each HE institution and a participating students’ choice of institution.

The nature of the research question imposes constraints on the quantitative research design. The location of institutions is, by and large, fixed and the individual decisions on whether and where to participate in HE are one-off, which explicitly limits us to a cross-sectional analysis. Hence our research design compares the decisions of different individuals who face different home-HE distance patterns according to where they lived (as a teenager) in relation to the spatial distribution of HE institutions. Using this variation in home location across individuals to infer behavioural traits is problematic because choice of residential location is determined by household and individual factors (like income) which may also determine HE decisions directly. We will thus need to rely on a regression-based strategy to control for observable factors – other than home-HE distance – that determine HE decisions and which may influence, or be influenced by, home location. To this end, our data allows us to link information on individuals’ HE decisions to detailed records on these students’ school test scores and qualifications, the school they attended, basic information on family background and Census data describing the characteristics of the neighbourhood in which their childhood home is located. We combine these data with information on the distances between each pupil’s home (at age
16) and each major HE institution in England. These data are described in more detail in the Data section below.

The main results presented are estimated elasticities of the probabilities of HE attendance (either participation at any institution, or choice of a specific institution) with respect to home-HE distance. The elasticities are estimated from individual-level logit models of HE participation, and conditional/multinomial models of institution choice. These models are well known, and we will not discuss them at length. The probability of a choice $j$ being made by individual $i$ ($P(i$ chooses $j$)) is expressed in terms of the natural logarithm of home-HE distance ($\ln d_{ij}$), vectors of observable characteristics of individual $i$ and choice $j$ ($x_i, z_j$) and estimable parameters ($\alpha, \beta, \gamma$) such that:

$$P(i$ chooses $j) = \exp(V_j) / \sum_k \exp(V_k) \text{ where } V_j = \ln d_{ij} \alpha + x'_i \beta_j + z'_j \gamma$$

(1)

The parameters are estimated by maximum likelihood methods using standard statistical software.

These logit models have an underlying theoretical economic justification in which $V_j$ represents a deterministic component to the individual’s utility or net benefits, associated with choice $j$ (the random utility model). However, in our application we will not be reporting the coefficients with this interpretation in mind. Our main concern is with the mean percentage change in attendance probability that is associated with a percentage change in home-HE distance, which can easily be derived from the data and estimated parameters. We estimate these elasticities for various student subgroups (by ethnicity, SES and income) either by estimating the models (1) separately by group, or by allowing for interactions between personal characteristics in the linear index $V_j$, for example by specifying

$$V_j = \ln d_{ij} \alpha_0 + \ln d_{ij} x'_i \alpha_1 + \ln d_{ij} z_j \alpha_2 + z'_j \gamma.$$

6 E.g. the elasticity of attendance at institution $j$ w.r.t log home-institutional distance for student $i$ is $\hat{\alpha}(1 - \hat{P}_j)$ where $\hat{P}_j$ is the predicted probability of $i$ attending $j$. 

- 7 -
An underlying assumption of these models is that individuals view alternatives as similar along those dimensions that are unobservable to the researcher, and hence not represented by variables in the probability defined in (1). For example, if a student is only making choices amongst top-rated research institutions in physics, then the distance to a low ranked institution not offering physics will not be relevant. It is necessary to fully control for all salient institutional characteristics that could influence choices, and allow for differences in preferences over these characteristics to fully overcome this problem. Alternative, more flexible formulations (multinomial probit, mixed logit) are infeasible given the number of students and pupils we have in our dataset. A ‘nested’ logit structure offers one way forward, but requires that we predefine groups of institutions that we consider as comparable alternatives e.g. top research rated universities offering physical sciences. However, individual decisions over subjects and institution types are likely to depend on the distances to institutions, and it is not straightforward to specify a sensible nested structure in terms of home-to-HE-institution distance\(^7\). In the absence of a feasible better alternative, we will use the standard conditional logit framework, but fully control, as far as is feasible, for the average preferences for each institution (using institution-specific dummy variables), and differences in these preferences across salient individual characteristics using interactions between institution, or institution type and personal characteristics.

Note that our specification of home-HE distance in logs means that the elasticity with respect to distance varies with the probability of attendance, but is otherwise constant at different distances, and implies that the marginal costs of distance are decreasing with distance, if \(V_{ij}\) is interpreted as a utility or net benefits term. This assumption has precedent in the transport and migration literature (e.g. for

\(^7\) There is no specific distance that relates to an institutional group that is not just some aggregation of the members of that group, which would make identification very tenuous.
“gravity” models of aggregated flows of migrants between regions.\(^8\) The setup outlined above for modelling institution choices has an analogous aggregate representation in which the flow of students between one residential zone and a given institution depends on the log of distance between that residential zone and institution. We will provide some non-parametric evidence that the chosen functional form for our choice models is appropriate.

Note that when we consider students’ choice of institution, we ask whether living closer to a particular university makes it more likely that a student attends that university, given where they live in relation to all other institutions. This approach, in part, takes account of differences in student background that are linked to residential location (i.e. considers whether distance of one institution relative to the rest affects an individual’s choice). In the case of the participation decision, we will be comparing potential university students who live close to institutions with those who live further away, and it will obviously be difficult to rule out unobserved individual differences between residents in dense locations (e.g. cities) and those elsewhere (e.g. rural) as explaining any association between HE distances and participation. However, because we are interested primarily in marginal participants for whom access to the closest universities is presumably the deciding factor, we have some scope to include geographical fixed effects (Local Authority or Region\(^9\)) to control for broader geographical differences in participation. In addition, we can estimate the influence of distance to nearest institutions, whilst controlling for average distance to all HE institutions.

3.1. Data

The empirical analysis is based on a composite data set that has been linked together from a number of administrative sources. The core of this dataset is a national cohort of English school pupils, \(^8\) See for example Fotheringham and O’Kelly (1989) on gravity models, or McGann (2005) on the possible relevance for New Economic Geography models.

\(^9\) These are administrative areas. There are 388 Local Authorities in England. There are 9 English regions.
sitting their end of compulsory schooling exams at age 16 in summer 2002, and whose demographic
details are recorded in England’s first Pupil Level Annual Census (PLASC). The PLASC is a national
pupil census that has been carried out by the Department of Children Schools and Families since 2002,
and contains information on school attended, pupil home address, ethnic group, gender, age and free
school meal eligibility (a basic indicator of low family income). Importantly, these data include pupils’
home address postcode, which pinpoints home location to the nearest 10 or so housing units.

In England, school children are tested periodically throughout their schooling, sitting nationally
assessed tests in a number of core subjects. It is possible to link the PLASC data on our cohort of
students to information on their academic achievement, recorded at age 11 and age 14, at age 16 when
pupils take General Certificates of Secondary Education or GCSE exams (academic) and National
Vocational Qualifications or NVQs (vocational) at the end of compulsory schooling, and at age 18
when those pupils who continue in education take Advanced Levels (A levels) or other post-16 higher
level qualifications. These attainment data, when combined with PLASC, provides a comprehensive
longitudinal record of each child’s secondary schooling.

For the first time, it is now possible to match this school information to additional data on each
individual’s subsequent decision to enrol (or not) in a higher education institution. These higher
education data are provided by the Higher Education Statistics Agency. As well as institutional details,
these data provide additional individual level information, including course type, subject choice, student
disabilities etc. We then link in other data sources to this composite data set, namely an indicator of
institution research quality based on the UK’s 2001 Research Assessment Exercise (the periodic

\[10 \text{ Akin to a zipcode.} \]

\[11 \text{ When pupils take the statutory Key Stage Assessments (part of the English National Curriculum assessment).} \]

\[12 \text{ The matching process was undertaken by the Department for Children, Schools and Families. Further information available on request.} \]
administrative review of the quality of research outputs produced by each university department),
degree drop out rates, and a description of university type. In England, universities fall into two broad
categories: ‘old’ traditional academically focused universities, and ‘new’ universities that were formerly
called polytechnics and were more vocationally oriented\(^{13}\). We then further sub-divide the ‘old’
university group into ‘Russell Group’, and ‘1994 group’ – both associations of research intensive and
high status universities – and ‘other’ institutions. We also separately indicate universities that are
specialist institutions offering a narrow range of subject choices, although we drop most specialist
institutions from our final dataset. From the full set of around 128 higher education institutions in
England we select the largest in terms of intake (accounting for 99% of enrolment), drop institutions
that enrol mainly postgraduate students and a few highly specialised institutions\(^{14}\). The full set of 96
institutions used in our analysis is listed in Appendix Table A1.

To this dataset, we further add geographic data from the 2001 population Census (at the smallest
Output Area level) to characterise pupil’s home neighbourhood, and aggregated economic data (wages,
housing costs) at “Travel to Work Area” level to characterise housing costs and future labour market
opportunities in the vicinity of HE institutions.

The central element in our empirical analysis of HE choice is the distance between a pupil’s
parental (or carer’s) home and each higher education institution. We compute the distances between
each home address of the 400,000 pupils in our data and each of the 96 HE institutions, via the rail
network. These home-HE distances are calculated using rail-network distances rather than straight line
distances, largely to avoid errors arising from infeasible shortcuts across river estuaries and the like\(^{15}\).

---

\(^{13}\) These two formerly different types of institution acquired the same description of “university” in 1992.

\(^{14}\) For example, the Royal Veterinary College, and the Open University which uses primarily distance-based learning.

\(^{15}\) We measure distances along the rail track that comprises the network, so these distances should be interpreted as the
shortest possible rail distance, rather than the actual distance that would be travelled. The correlation between straight line
distances and rail network distances is very high (around 0.99) and both give almost identical results. The chosen
mechanism for calculating distances is therefore unlikely to be an source of error (akin to Combes and Lafourcade (2005)’s
findings for the French transport network).
In summary, this complex combined dataset provides a new and unique opportunity to study higher education choices at the pupil level, combining institutional information with detailed information on students’ academic achievements at school, demographic background and details of home location. In the next section we move on to present our empirical results.

4. Empirical results and discussion: participation

4.1. Mapping university accessibility and participation

A first step is to consider the spatial distribution of institutions and the geographical accessibility of these institutions to pupils’ home residences. We focus on the set of 96 major higher education institutions listed in Appendix A1 (described in Section 3.1).

Figure 1 maps key indicators of the geographical accessibility of HE institutions from pupil homes. In the first panel (A) we show how distance to the nearest three institutions (regardless of size of institution) varies across England. In the second panel (B) we show mean distance to all HE institutions. In Panel C, we estimate geographical accessibility based on the number of first degree places available within 100km of each pupil residence. In both cases the HE accessibility indicators are calculated for each pupil residential postcode, and the maps are created by GIS interpolation of these data on to a 1km square raster.

As we might expect, Panel A suggests that proximity to HE institutions is higher in urban areas. The white areas indicating 3 HEs within 20 km tend to delineate the main urban centres in England. Having said that, very few areas have very poor geographical access to HE and residents in by far the largest part of the land area of England have 3 institutions within 80 kilometres by rail (50 miles). Only in a few peripheral rural areas around The Wash, north Norfolk, Lincolnshire to the East, north Devon and Somerset to the South West, and Cumbria in the far north west, are the nearest 3 institutions on average over 100km away (and in the North this figure may be distorted by the fact that we do not include Scotland’s HE sector in the analysis). Panel, B, shows for comparison, how far pupils are away
on average from all institutions and highlights the fairly obvious point that living centrally in England gives students closer access to the full set of institutions. The map in Panels A and B ignore differences in institution size, so perhaps paint an unfair picture of the accessibility of HE places. Panel C on the other hand maps the distribution of university places. This makes it transparent that the centre of mass of higher education access in England is located in London and the Midlands and central North West, where distances to large urban universities are short and up to 100,000 first degree places are to be found within 100km of a pupil’s home. This pattern is hardly surprising, given that the population of England is predominantly urban in location, and university supply has evolved to meet demand. However, a large number of places are accessible within 100km even in the peripheral areas of England’s coasts and borders.

Figure 2 turns away from considerations of HE accessibility, and looks at the spatial distribution of participation. This map gives a feel for the geographical distribution of HE participation, but it is not sensible to try infer any relationship between the accessibility of HE and participation from simple visual comparisons. On the one hand there is a temptation to read the relatively low participation rates in areas peripheral to HE (particularly the South West, and eastern coastal areas) as symptomatic of poor institutional accessibility. On the other hand, accessible urban areas (the South East, central midlands and North West) have a mixture of high and low participation pockets. These patterns are likely to depend to a large extent on the residential sorting of households of different types (incomes, academic achievements etc.) into different geographical areas, rather than any ‘causal’ linkages. We will consider the links between accessibility and participation in the statistical analysis that follows.

4.2. Descriptive statistics

Table 1 shows 4 indices of the geographical accessibility of HE institutions, summarised by pupil characteristics and by HE participation status. Figures are means, with standard deviations in parentheses. Looking at the White British ethnic group first, we can see that individuals live, on average, quite close to HE. The average distance to the nearest three HE institutions is around 36 km.
(22 miles), but there is almost no difference in the mean between non-participant and participant groups. Participating pupils, do, however come from homes that are slightly closer (by about 1%) to all HEs on average, and where there are 1-1.5 thousand more first degree places within 100km. One factor that could make a difference to the participation decision is the quality of local institutions, so we consider too the distances to ‘high quality’ institutions categorised as those receiving a top 20% Research Assessment Exercise grading averaged across all academic departments. Again, participants and non-participants are not strikingly different in terms of the magnitude of the distance from high quality institutions.

Non-White British students are predominantly urban and live, on average, much closer to HE and closer to high quality research institutions. Their average distance to three institutions is under half that for White British students. Otherwise, the pattern across participants and non-participants for these ethnic groups is broadly similar to that for White British, with only small differences between participant and non-participant pupils in terms of HE proximity, and in some cases (e.g. Black, and ‘other’ ethnic groups – which includes Chinese) participants have poorer HE access. The most notable gaps occur for access to high quality institutions (top 20% RAE) – for example, Pakistani and Indian females who participate are around 4-4.5 km (9%) closer to their nearest three, high RAE-rated institutions than those who do not participate. The magnitudes of any differences between participants and non-participants are slightly stronger for women than men.

In the last two columns of Table 1, we look only at White British students, and investigate associations with income using an indicator of free school meal entitlement for the pupil when s/he was aged 16. This indicator is the best individual level measure of income disadvantage available for our non-participant sample. Although less marked than for ethnic groups, low income students tend to be close to their nearest university (again reflecting their relatively higher rate of urbanisation) and closer to their nearest high quality institution, although White FSM teenagers tend to live marginally further away on average from all universities on average.
These simple comparisons suggest no compelling grounds for suspecting that accessibility plays a key role in influencing HE participation decisions for any of these student groups, although we have not yet controlled for family socioeconomic characteristics that may lead those with high HE participation propensities to live relatively near to HE institutions (e.g. in cities) or equally that may lead poorer students with potentially lower HE participation propensities to live in urban areas. In the next section we extend these findings using logit regression models of participation to control for other observable student and area characteristics.

4.3. Regression models of HE participation

Table 2 reports the elasticity of participation with respect to three HE access indicators: distance to the nearest HE institution, distance to the next two, and distance to those remaining (controlling for characteristics of students, schools and neighbourhoods in individual level logit models). This elasticity gives the relative percentage change in participation probability with respect to a one percent relative increase in home-HE distance. The coefficients on the control variables (listed in the table notes) are not reported for reasons of space and clarity. Note that the specification estimates the relationship between each distance variable, conditional on the others. Hence, for example, the elasticity with respect to nearest HE corresponds to a conceptual experiment which moves the nearest institution closer whilst keeping all the rest where they are. We also report the results of two statistical tests: (i) that all the coefficients on the log-distance variables are zero and (ii) that they sum to zero, which would imply that a 1% change in mean distance to the nearest, next two and remaining institutions has no effect on participation.

First, in Table 2 we estimate the model by ethnicity. The results for White British teenagers (Column 1) suggest that participation probabilities decrease with distance to closest institutions, for both females and males, and these are strongly statistically significant effects. In contrast, the positive coefficient on the variable measuring the mean distance to institutions beyond the nearest three suggests that it is students who are more generally remote from HE that participate more, but we prefer to treat
this variable as a control for unobservable factors that affect participation and residential location, rather than as an indicator of the likely effects of HE accessibility on participation\textsuperscript{16}. The coefficients on closest institutions can, arguably, be given a more causal interpretation. In these cases, the effects are statistically significant, but magnitudes are small, and the implications of quite large relative changes in distance seem fairly trivial. For White British females, doubling of distance to the nearest institution would reduce the probability of participation by about only 4.5\% in relative terms – reducing the probability of participation at the mean from 28.4\% to 27.1\%. For males, the effect is smaller still. In the context of other work (Crawford et al. 2008), which has found a strong relationship between prior attainment and HE participation, these geographical effects seem particularly trivial.

Across the different ethnic minority groups, the coefficients are of a similar order of magnitude for the first two distance measures (although distance is not always significant for all groups). It is hard to spot any systematic relationship with gender or ethnic group. There is no evidence here that ethnic minority girls, a group that has often been identified as potentially sensitive to geographical barriers to HE, are any less likely to participate than white girls because they live far away from their nearest institutions. In fact, the coefficients for Bangladeshi students are all individually and jointly insignificant. However, Black, Bangladeshi and Indian pupils who live further away from institutions beyond the nearest three are much less likely to participate; which probably reflects that those families with a high propensity to participate, choose to live in places that have large numbers of institutions close by, e.g. in the centre of England’s major cities.

The last Column of Table 2 presents analogous results for low income families, that is, White British students entitled to FSM. The distance elasticity for FSM pupils is generally larger than for the White British group as a whole (Column 1). Overall though, the patterns are qualitatively similar and provide only weak evidence that for low-income students distance is a more important barrier to

\textsuperscript{16} The magnitude and sign of this coefficient varies tremendously across the different ethnic minority groups, which may also support this interpretation given the different geographical clustering of different ethnic minority groups.
participation. For example, doubling the distance to the nearest institution would reduce the probability of participation by 5.7% for FSM girls, which would reduce FSM participation from 7.6% to 7.2%, at the sample mean.

These findings are robust to measuring distance to nearest or nearest three ‘high quality’ research institutions. Also, inclusion of Local-Authority-of-residence dummy variables in these models tends to render all distance-to-HE variables insignificant, or unsystematically signed. This suggests that marginal variation in HE distance between pupils in a given Local Authority is unrelated to the decision to participate, and also means we cannot be confident that the small distance effects shown are not simply due to unobserved differences between the pupils living in different Local Authorities, or even local policy or administrative factors. On balance, the regression results support the view that geography has a very limited role to play in the HE participation decision, in line with the findings in Sa et al. (2004) for the Netherlands.

5. Empirical results and discussion: institution choice

5.1. Visualisation and descriptive statistics

In this section we move on to consider the role of home-institution distance on choice of institution.

To motivate the analysis, Figure 3 presents the geographical picture, showing the distance (by rail network) between home and the actual institution attended for students in our dataset. The map is very similar to the first panel of Figure 1 (the map showing the distance to the nearest three institutions) rather than the second panel (which shows the average distance to all universities). This immediately hints that there is a strong tendency for students to choose amongst the institutions that are closest to home. As we might expect, it is students in the peripheral areas of the South West, East Anglia and the North that make the longest migrations to attend university.
Table 3 presents some descriptive statistics for distance to HE attended, split by gender and ethnicity and by parents’ occupational group and FSM status (for White British students). Average distance to institution attended (Table 3) is considerably greater than the average distance to the nearest three institutions (Table 1), but considerably less than the average distance to all institutions, for all groups. For White British females the distance attended is 133.5 km, which is 3.7 times the distance to the nearest three institutions and 55% of the mean distance to all institutions. For men, the distance is 141.5 km, and the corresponding ratios are 3.9 and 59% respectively, indicating a tendency for men to migrate or travel slightly further to HE.

Across the ethnic groups these ratios differ widely. Black students attend institutions at a shorter distance (79.6 km for females, 74.7 km for males), but this is because they live in places close to universities. These distances are over five times the distance to their nearest three, and just over 40% of the average distance to all institutions, suggesting that the average Black student tends to pick institution that are relatively far away from home. For Pakistani and Bangladeshi students the situation is reversed and they pick relatively nearby institutions. Males and females of Indian ethnic group travel or migrate far, considering that they have institutions close by.

For White British students, low income students who are entitled to FSM (middle panel) travel less far than students from non-FSM backgrounds, both in absolute terms (about 30% less) and relative to average institutional distances. The pattern for parental occupational groups also suggests an association between distance travelled and parental occupational status, with children of Professional parents attending institutions furthest away.

None of these descriptive findings indicates whether the differences in travel/migration distances are due to constraints on travel behaviour (e.g. by income or cultural background) or due to institutional preferences for these different groups. For example, a student may prefer a city university to a campus-based one. If this preference is prevalent amongst students from low income backgrounds, and low-income families tend to reside in cities, then we will find that low income students attend nearby
institutions (those in their home city), even though the distance away from home is not the factor driving their choice. In the next section we extend the institutional choice analysis to control for institutional preferences and background characteristics, particularly students’ prior attainment.

5.2. Regression models of institution choice

The conditional logit framework outlined in Section 3 provides a method for estimating the effect of distance on attendance probabilities, whilst controlling for personal and institutional factors. These estimates are based on the ‘dyadic’ pupil-institution dataset in which each observation is a pupil-institution pair (i.e. with all possible pupil–institution combinations in the sample). Home-HE distance clearly varies within HE institutions between pupils, but also within pupils between all possible HE institutions.

Institutional characteristics (institution type, RAE scores etc.) can enter the models directly, but the alternative and more general specification is to include a full set of institutional dummy variables, which account for general differences in the attractiveness of each institution to students in the estimation sample. In addition, institutional characteristics can be interacted with pupil characteristics (including home-HE distance). The effect of individual characteristics on institution participation cannot be measured unless individual characteristics are interacted with either: a) indicators of the specific institutions in question (e.g. the effect of pupil GCSE scores on the probability of attending the University of Leeds); b) indicators of given groups of institutions (e.g. the effect of pupil GCSE scores on the probability of attending a top RAE-ranked university); or c) with some other characteristic that varies across institutions for a given pupil (e.g. interacting GCSE score with home-institution distance).

We also undertook a non-parametric kernel regression, of an 0-1 indicator of institution attendance on home-HE distance, to show how the probability of attending an institution changes with the distance between a student’s family home and institution. These analyses suggest that the parametric conditional logit specification provides a reasonable approximation to observed choice behaviour (see also Appendix Figure A1, where we plot the predictions from the logit model against the non-parametric estimates for an example population group (White British females)).
shows the association between GCSE scores and the propensity to attend a university far away from home – because home-HE distance varies by institution for a given pupil).

Given that we have nearly 100 institutions and some 25 personal, school and neighbourhood characteristics, option a) is computationally infeasible because it would require estimation of 2500 parameters. We do however, report estimates in which we allow a limited set of personal characteristics to influence specific institutional choices, in which we interact pupil characteristics with key institutional characteristics, and in which we interact key personal and institutional characteristics with home-institution distance. We also estimate our models separately by ethnic, income and occupational groups.

These central results, split by gender, are provided in Table 4, for ethnic groups, and Table 5 by parental occupational groups and FSM status. The top panel of each table reports results for females and the bottom panel for males, and each pair of columns reports elasticity estimates and the t-statistic for the underlying coefficient on which it is based\(^{18}\). Each pair of cells is an estimate from a separate regression, and we report results for four specifications for each group of students. Firstly, we report an unconditional elasticity, with no student or institution control variables. Below that, we report the elasticity conditional on institution-specific dummies to control for all institution specific factors that affect attendance in the same way for all students in the estimation sub-sample. In the third row, we interact these institution dummies with student GCSE point\(^{19}\) scores, to allow for the fact that student

\(^{18}\) Given that there are around 96 institutions, the probability of choosing any one at random is 1.04%. The elasticity is calculated by multiplying the estimated coefficient on log distance by \((1\text{-probability of attendance})\), hence the average elasticity is not very different from the underlying coefficient in magnitude.

\(^{19}\) We use GCSE (age-16) point scores rather than A-Level (age-18) scores, because the former provide the finest possible measure of student prior achievement/ability, predate students’ choice of HE institution, and are non-missing for a higher proportion of students in our source data A-Level and GCSE scores are, in any case, highly correlated. We do however include A-Level scores interacted with institution characteristics.
preferences and university admission criteria will affect the probability of attendance at different institutions. In this specification we also interact a set of institution characteristics (institution RAE 2001 score, and dummies for institution type) with a limited set of pupil characteristics (age, English Additional Language, FSM, school proportion on FSM, A-Level scores, occupational group in ethnic models or occupational group in the ethnic models).

The specification in the fourth row in each panel allows for interactions between home-institution distance and personal and institution characteristics. In this case, the attendance-distance elasticity varies by personal and institution characteristics. Thus the elasticity reported in the table corresponds to a specific baseline group of pupils who are non-FSM, English first language, with parents in non-managerial/professional/administrative/skilled occupations, from Community non-selective schools, enrolled on a business, creative, or administrative degree, at a ‘new’ university that is not a specialist institution. This baseline elasticity is estimated at the mean of continuous variables (GCSE scores, RAE rating, neighbourhood characteristics etc). A more complete set of results for these specifications is shown in Appendix Tables A2 and A3.

Looking down the column for White British females and males, the first three elasticity estimates are all very close to minus one and the coefficients very precisely estimated. Evidently, the estimates are very insensitive to whether or not we control for pupil or institution characteristics. If given a causal interpretation, this means that a doubling of the distance between a student’s parental home and an institution picked at random, halves the probability of the student attending that institution. To an approximation, from Table 1, this means that a White British student is over 6.5 times more likely to attend an institution located at the average distance of their nearest three, than an institution at the average distance of all institutions \((241/36)\) from Table 1). Gender has no bearing on this relationship for White British students.

For the other ethnic groups in Table 4, we see some important differences in the first three rows. Distance imposes less of a cost for Black students, especially Black females, and a considerably higher
cost for Bangladeshi and Pakistani students, especially females. Indian, ‘other white’ and other ethnic
groups exhibit no difference from White British students in the sensitivity of their choices to home-HE
distance. The higher elasticity for Bangladeshi and Pakistani women is in line with anecdotal,
qualitative and earlier statistical evidence that students in these groups tend to stay close to home. The
figure of -1.5 for these students implies that doubling the distance between home and institution reduces
the probability of attending that institution by a factor of 35%, (= 2^{-1.5} ) and means that a Bangladeshi
woman is over 50 times more likely to attend one of their nearest three institutions than an institution at
the average distance of all institutions ((200/14)^{1.5} from Table 1). Even if Bangladeshi women faced
the same distribution of HE distances as White British females, they would be over 2.5 times more
likely to attend one of their nearest institutions than their White British counterparts ((241/36)^{1.5-1}).
This feature is less marked, but still present, for Bangladeshi and Pakistani men.

The 4th row in each panel allows for distance-characteristics interactions, to better compare pupils
from different ethnic groups on a like-for-like basis, in particular allowing for prior achievement.. The
elasticity changes between rows 3 and 4, because the estimate no longer corresponds to the mean in the
sub-sample, but to the baseline group of pupils (White British, non-FSM, etc.) defined above. This
specification shows that at least part of the difference in attendance-distance elasticities that we
observed across ethnic groups is attributable to factors other than ethnic group. For example, the
institution choices made by baseline White British girls and baseline Bangladeshi girls show a similar
response to distance now that we compare students who have family and educational backgrounds that
are alike. On the other hand, comparable Black students still respond less to distance than White British
students, and Bangladeshi boys and Pakistani students of both genders are more sensitive to distance.
Many factors turn out to be more important than ethnicity in determining willingness to travel. For
example, the elasticity is -1.5 for White British students at the bottom of the achievement distribution (2
s.d. below the mean in GCSEs and A-Levels – see Appendix Table A2 and A3) implying that low-
achieving students are less likely to travel far to HE.
Institutional factors measuring aspects of university quality can prove strong attractors that mitigate the deterrent effects of distance: a top RAE rating (2.s.d. above the mean) reduces the distance elasticity of students by about 15-20 percentage points for White British students, whilst low drop out rates are strongly associated with greater willingness to travel. However, institutional factors such as high-RAE ratings and low drop out rates do little to mitigate the deterrent effect of distance in institution choice for Bangladeshi and Pakistani minority groups (see Appendix Table A1 and A2), and additionally for these groups, qualifications are not so strongly linked to willingness to travel.

We now turn attention to differences by parental occupational and by FSM status, for White British students only. These results appear in Table 5, which has the same structure as Table 4. Students from Managerial and Administrative parental backgrounds tend to exhibit a university choice behaviour that is representative of the White British group as whole with a distance-elasticity close to minus one. The elasticities are smaller (implying greater willingness to travel) for Professionals and larger for the lower ranked occupational groups (implying less willingness to travel) even controlling for institutions and their interaction with personal characteristics. There are clearly fairly marked differences in travel to HE behaviour across income groups – for example, students from FSM backgrounds would be nearly 12 times more likely to attend an institution at the average distance of the nearest three, than at the average distance of all institutions. We are, given our data, unable to determine whether these differences are causally related to income, or whether they are due to other factors embedded in these occupation and income definitions (expectations, cultural traditions, norms etc.).

Note however, that when we allow for interactions between distance and other background and educational characteristics in the 4th row of each panel in Table 5, the occupational background differences become far less marked – particularly for boys. For example, after allowing for the large differences in prior achievement of FSM and non-FSM students, FSM students appear to be more similar to non-FSM students in terms of their sensitivity to distance. Distance is largely irrelevant to White British students from any parental occupational background if students have high qualifications,
enter a university with a top RAE-rating and a low drop out rate and take a gap year (from Appendix Table A4-A5). This is what we might expect of ‘traditional’ British university student behaviour.

An important point emerges from these regressions: home-institution distance is the single most important factor determining institution choice. Compare for example, distance with GCSE scores (our preferred measure of ability and prior achievement). GCSE scores have a large and significant association with institution choice (as we would expect) and a one-tenth of one-standard deviation decrease in GCSE score reduces the probability of attending the most prestigious institutions by about 7 percentage points, due presumably to a combination of institutional entrance requirements and student preferences. But in terms of its effect on institution choice, the costs imposed by this fall in GCSE points is equivalent to a mere 7% increase in home-HE distance.

One important health warning must accompany these results: we cannot rule out empirically the possibility that this attendance-distance relationship arises because parents choose where to live on the basis of which university they expect their children to attend, rather than university choices being made conditional on home location (decided upon for other reasons). Either way, the estimated attendance-distance elasticity reflects the costs imposed by distance, although the policy implications are very different e.g. policy to improve the geographical accessibility of HE will be completely ineffective in changing patterns of HE attendance if we are measuring a selection effect, whereby parents whose children are more likely to enrol in high quality institutions choose to locate near to such institutions. Our only defence in this respect is that introspection suggests that university choice is not an important factor in residential choice, and to our knowledge no previous literature has raised this possibility (this is in stark contrast to the case of school choice, where how close a family lives to a school affects the probability of admission). Moreover, from the lack of association between participation and HE home-distance (Tables 1 and 2) it seems unlikely that families choose homes location according to whether they intend their children to participate in HE, so it seems most unlikely that residential sorting matters for specific institutional choices.
5.3. Implications for patterns of HE institution choice

Are these differences across ethnic and income groups of policy concern? On the one hand the estimated elasticities are the result of preferences and rational economic behaviour, since it makes sense to minimise the financial or psychic costs of travelling to university if perfectly good local options are available. On the other hand, if distance has such a strong impact on institutional choice it does lay open the possibility that choice is inherently restricted by where a person lives. In line with our previous work on the representation of ethnic minority students in high status institutions (Chowdry et al. 2008), the conjecture that distance acts as a constraint on access to high quality universities amongst ethnic minorities does not bear close scrutiny. As is seen from Table 1, ethnic minorities generally live closer to high quality institutions than the average White British student. If we look at Table 6 – which tabulates the proportion of students scoring in the top 20% of the national distribution in GCSEs who go to top 20% ranked research universities (by 2001 RAE score), we see that high-achieving pupils in all ethnic minority groups\textsuperscript{20} are over-represented in top research institutions relative to White British. On this evidence, it doesn’t look likely that distance to top institutions drives differences in the types of institution chosen by different ethnic groups.

Can distance to HE explain any other patterns of institution attendance in Table 6? A striking feature for all ethnic, occupation and income groups is that high-achieving women are under-represented in top research institutions. Distance could be a contributory factor if a) women live further away from top research institutions than men or b) if distance acts as a stronger deterrent to attendance than it does for men. The first conjecture is both theoretically unlikely, and empirically unsupported (see Table 1). On the second conjecture, quite small gender elasticity differences in the range shown in Tables 4 and 5 could be responsible for part of the gender gap in attendance rates at top research institutions: high achieving male attendance at top ranked institutions is 15% higher than female attendance for women.

\textsuperscript{20} Apart from Pakistani girls, where the difference is not significant relative to non Pakistani girls.
attendance. For students choosing amongst their nearest 3 institutions, it can be shown that a 17 percentage point gap between the attendance-distance elasticity of men and women is necessary to explain the gap in participation between men and women in top RAE-ranked institutions. For some ethnic groups, the gender elasticity differences could well explain this gap – e.g. for Bangladeshi and Pakistani students. However, for White British students and other ethnic groups the gender differences in distance elasticities exhibited in tables 4 and 5 are, for the most part, well below what would be necessary to generate the differential attendance patterns in terms of quality HE (it makes sense to focus on the 3rd row in each panel in Table 4, because the 4th row controls for the differences in willingness to travel to institutions of different types and quality).

Similar reasoning points to a much stronger role for parental occupation and income differences in willingness to travel to university. Table 6 shows that there are substantial differences across occupational groups in students’ attendance at top research institutions, even for high achievers at GCSE level. High achieving FSM pupils have a particularly low probability – only 48%. Looking back at Table 5, the elasticity differences across occupation and income groups can well explain these attendance differences. For example, a high-qualified woman from a professional parental background is 24 percentage points more likely to attend a top-RAE institution than a woman from a skilled-trades background (Table 6). The elasticity difference between professional background and skilled-trades background women (Table 5) is as high as 42 percentage points, which can more than explain this difference in attendance probability.

\[ \text{Assume men and women live the same distance from institutions and have an equal probability of attending the average of their nearest three institutions at distance 36km. The mean distance of the nearest 3 top RAE institutions is 82km. Thus, the relative probability of top-RAE attendance is } \left( \frac{82}{36} \right)^e \text{ where } e \text{ is the attendance-distance elasticity. Some simple manipulations show that the difference in elasticity necessary to make men 15% more likely to attend top RAE institutions because of distance, is } \frac{\ln(1.15)}{\ln\left( \frac{82}{36} \right)} = 0.17. \]
In summary, differences in the ‘costs’ imposed by distance can provide an empirically credible explanation for differences in the quality of HE institution attended, for White British students from different parental occupation and income backgrounds, although not for differences amongst ethnic groups, or by gender.

6. Summary and concluding remarks

This paper has provided empirical evidence on the association between home-institution distance and the probability of college participation, and of choosing a particular institution. Our results suggest that participating and non-participating pupils’ homes are similarly distributed in relation to the location of HE institutions in England, and there is, at most, a very weak link between home-HE distance and the decision to participate in HE. Geographical proximity of HE is not an important factor relative to others, particularly early academic achievement (Chowdry et al. 2008), in determining higher educational participation. This would imply very little scope for policy to widen participation through increased geographical accessibility, either through manipulation of institution location, by transport improvements, or other means. As a corollary, it seems that university location plays no role in forming geographical disparities in basic human capital accumulation through the channel of enrolment, although the migration decisions of students after leaving HE may still be an important factor in this respect (Faggian et al 2006. 2007). Given that in England most types of student live less than 40km away from their nearest three universities (Table 1), the fact that geography is not a more critical factor is unsurprising.

In contrast, home-institution distance is strongly linked to institution choice. In fact, distance emerges as the most important general factor associated with institution choice. Consequently, university intakes are, on average, skewed disproportionately towards those students whose parents live relatively close-by. This in itself is likely to explain the apparent over-representation of some ethnic minority groups in some inner-urban universities. Even so, some ethnic groups – especially Bangladesh and Pakistani girls – appear to be considerably more sensitive than others to home-HE distance when it
comes to institution choice, and possible reasons for this have been documented elsewhere (for example, Mac an Ghaill and Haywood, 2005). However, once we take account of differences in students’ prior achievement and other personal characteristics, and also given that students in these particular ethnic minority/ gender groups live relatively near to high quality institutions, there does not appear to be any correspondence between this greater distance sensitivity and differences in the quality of institution attended.

White British students from different parental occupation and income backgrounds differ too in the sensitivity of their choices to home-HE distance, with (to a rough approximation), the sensitivity increasing as income and occupational ‘status’ decreases. Moreover, students from low income/status parental backgrounds have a low probability of attending high research-quality institutions relative to their equally qualified peers from better-off backgrounds. This pattern is consistent with the facts that a) high quality HE institutions are further away on average for these students, and that b) poorer students are more sensitive to distance when it comes to institution choice. The higher distance sensitivity of poor students may be the result of cost barriers. Such students may need to reduce the costs of attending university by choosing a proximate institution, even if that is lower quality. The findings therefore offer some support for the idea that targeting the accessibility of higher-quality HE institutions could increase uptake of high quality HE places amongst suitably qualified students from lower-occupational status backgrounds. Such policies might include action to reduce the role of distance (distance learning) but also policies to encourage higher status institutions to undertake outreach activities further afield. We find no evidence to suggest that such a policy needs be gender or ethnically targeted.

One further important spatial implication from this work is that the type and quality of HE education in which students enroll is in part governed by the type and quality of local institutions, which in turn partly determines the skill composition of the local population. Given this, the local mix of HE institution types and quality could have a strong bearing on the quality and composition of the local human capital stock.
7. References


