Department for Education

Secondary school choice and selection

Insights from new national preferences data

Research report

August 2018

Dr Matthew Weldon – Department of Economics, Lancaster University

GSR GOVERNMENT SOCIAL RESEARCH
Social Science in Government

Contents

List of figures

List of tables

1 Executive Summary

This report exploits newly-available data from the secondary schools admissions process to shed light on the ways in which parents' decision-making when choosing schools, and experience of gaining admission to chosen schools, differs in different English cities, and for different demographic groups within those cities.

We show how the data can be used to inform two current themes in education policy:

Social mobility and the London effect

Secondary schools in the capital have been shown to outperform schools in the rest of England, including other large cities. Children of disadvantaged families appear to do particularly well, compared to similar children outside London.

Among other hypotheses, it has been suggested that the school choices of London's parents may play a role in explaining the effect: if London's parents are more demanding of quality, and more willing to go beyond their local school to seek out high-performing schools, this might create competitive market conditions that lead to school improvement across the board. Of special interest is the choice behaviour of disadvantaged families in London, as the relative success of children in this group may be related to the decisions their parents make about their children's education.

Comparing London, Manchester, Birmingham, and towns in the Pennines region of Northern England, the report asks whether geographical variations in choice behaviour can help to explain the apparent variation in the success of education markets, the 'London effect', and the existence of educational 'cold spots'. The report also investigates whether geographical differences remain after accounting for demographic variation.

School Admissions and equality of access

One of the main limitations to choice is constrained capacity at good schools. At popular schools the number of applications often far exceeds the number of places at the school, and some oversubscription criteria must be applied to determine who is allocated a place. At the same time, it has been observed that the intakes of some types of popular schools tend to be more advantaged, more able, and less ethnically-diverse than their local neighbourhoods.

Competing explanations of this apparent sorting invoke either choices (i.e. parents self-sort by choosing schools differently), residential sorting (i.e. house prices around popular schools drive out less-well-off families) or selection (i.e. admissions policies affect different groups differently).

Hitherto, distinguishing between these competing explanations has been difficult. The new data allows us to control for both choices and residential location, to isolate the residual effect of admissions policies on the demographic sorting of secondary schools.

1.1 Methods

In addressing these questions, the report focusses on variation along four dimensions:

- Geography;
- Disadvantage (Pupil Premium status);
- Child's prior academic attainment at primary school;
- Ethnic group.

We look for evidence of geographical variation (e.g. a London effect) that persists after controlling for the other dimensions.

Three methods are used in the report. First, **descriptive statistics** tabulate the proportion of children obtaining their preferred school, and the proportions ranking different numbers of schools, by location, Pupil Premium, attainment and ethnicity. The tables are used to explore the choice behaviour and relative admissions success of different groups.

Second, a **statistical choice model** is used to estimate the relative weighting of academic performance (relative to convenience and other factors) in parental decision-making, and parents' willingness to seek out higher-performing schools that may be further from the family home. This analysis is also broken down by location, Pupil Premium, attainment and ethnicity.

Finally, we implement a **matched case-control design** to examine the probability of being admitted to a school, having applied for that school. By matching on location, this design eliminates the 'selection by mortgage' element of admissions, to isolate the ways that school admissions practices differentially effect different demographic groups.

1.2 Results

We find evidence of demographic variation in both decision-making and outcomes from the admissions process, with particularly striking variation between white and minority ethnic groups. Some headline findings include:

- In England 93% of white British families obtain their most-preferred school, compared to only 73% of black families and 75% of South Asian[1] families.

[1] Indian, Pakistani, Bangladeshi and mixed ethnicities containing any of these.

- The difference between Pupil Premium and non-Pupil Premium families is more modest: 86% of Pupil Premium eligible families obtain their first choice, compared to 89% of non-Pupil Premium families.
- In London, Manchester and Birmingham we find that families of minority ethnicities appear to weigh academic performance more highly than white families in choosing a school. For example, black and South Asian parents on average travelled more than twice as far as white parents for a similar improvement in the school's average test scores (measured by the percentage of pupils achieving 5+ A*-C at GCSE), although this may be partly explained by residential sorting. In the Pennines sample, in contrast, families of minority ethnicities appeared to travel a shorter distance, on average, for test score improvements than white families.
- Parents do not always use all of the preferences they are allowed, but families of minority ethnicities rank more schools on average than white families.
- Black families are 68% *more* likely to choose a Church[2] school than white families, yet they are significantly *less* likely to be admitted to a Church school than a similar white family living nearby. If a white child and a black child apply for a single remaining seat at a Church school in London, the black child is less than half as likely to be admitted. We find similar results for other ethnic groups in London and other cities.
- Likewise, a Pupil Premium-eligible child is significantly *less* likely to be admitted into a Church school she applies to, than a similar non-Pupil Premium child living nearby.

After disaggregating variation by demographics, and particularly ethnicity, subtle differences between London and the other cities remain, but it is the ethnic variation in both choice behaviour and access to schools that stands out.

1.3 Conclusions

On the London effect, we find that differences between parental decision-making in the Capital and elsewhere can largely be explained by London's diverse ethnic mix. The Capital has a far larger proportion of families belonging to minority ethnic groups than Manchester, Birmingham or other areas, and there appear to be differences in the ways parents of different ethnic groups choose schools. We do not know whether these differences are related to differences in preferences for quality, or differences in access to schools, or even differences in residential decision-making. However, whatever the causes, these differences do affect aggregate parental demand for academic standards. Following a 'quasi-market' argument, the greater demand for academic standards in

[2] Church schools are schools having either a Roman Catholic or Church of England denomination.

London might give schools stronger incentives to compete on quality, which might plausibly be one ingredient in explaining the relative success of London's schools.

The report cautions against interpreting choices without considering strategic pressures on parental decision-making, such as the possibility that parents may feel it is unwise to list schools where the child has little chance of admission. Such strategic incentives distort the picture that preferences data gives us about parental preferences, and they also tend to exaggerate headline estimates of parental satisfaction with their children's allocation (i.e. the percentage allocated to their 'first choice' school).

On admissions practices, the report finds evidence that children of minority ethnicities, and disadvantaged (Pupil Premium) children, are less likely to gain a place at oversubscribed own admissions authority[3] schools, and Church schools in particular. For minority ethnic groups such as black families, who are much more likely to choose Church schools than white families, the selection effect may have a noticeable effect on parents' satisfaction with the admissions process.

The causes of these systematic differences in admission probabilities are not clear. One possibility is that families from different backgrounds have different expectations with respect to the chances of admission to own-admissions authority schools, and Church schools in particular. For example, it may be that non-white or disadvantaged families underestimate the level of competition for Church school places, and consequently do not place as much emphasis on fulfilling religious or other requirements before applying.

The results do not imply that Church schools are cream-skimming pupils, nor do they provide evidence that admissions oversubscription criteria are not being applied consistently. However, the results provide strong evidence that differences in the composition of Church schools and other types of school cannot be explained entirely by parents' preferences and are, at least in part, due to admissions constraints.

[3] In this report, 'autonomous' schools are defined as those schools that act as their own admissions authorities. This includes academies, free schools, foundation schools and voluntary aided schools.

2 Introduction

In 2013 the Department began to collect individualised data on parents' ranked preferences from local authorities' admissions processes. The data has been linked to the National Pupil Database, and the linked dataset constitutes a unique opportunity to study the ways that parents choose schools, and parents' diverse experiences of the admissions process.

This report has been commissioned to demonstrate the potential of the linked data to inform key areas of the Department's policy agenda. We focus on two timely themes: social mobility and the London effect; and the effect of school admissions arrangements on access for different groups of families.

The report focusses on admissions to secondary schools in 2013, although the linked data currently extends to primary and secondary school admissions in 2013 and 2014.

2.1 Social mobility and the London effect

In recent years there has been mounting evidence that both academic attainment and progress in London are higher than in most other parts of England. The difference in performance is particularly stark amongst disadvantaged pupils in London, who outperform non-disadvantaged pupils in some other parts of the country. In the light of this finding, which has been dubbed 'the London effect' (Cook, 2013) there have been several accounts put forward to attempt to explain the effect.

The first set of accounts emphasised the importance of deliberate policy initiatives put in place during the early 2000's, such as the 'London Challenge' and 'Teach First' initiatives and the academies programme (Baars *et al.*, 2014). Other studies suggest that the process of improvement in London began before these policies were implemented, and credit a more gradual systemic improvement in standards (Blanden *et al.*, 2015). Burgess (2014) finds that much of the 'London effect' can be explained by the relatively diverse ethnic composition of the Capital, and argues that demographics, not policies, play a key role in explaining London's success.

The report contributes to this literature by examining the extent to which the 'London effect' and the comparative performance in other English cities, can be explained by the structure of education markets. It has been hypothesised that a more effective market structure, facilitated by market-based reforms such as the academies programme, may play a role in the Capital's success. That is, differences in the diversity of options available to parents, and the way parents choose between those options, may serve to drive a competition between schools that drives up standards, and somehow does so more effectively in London than in other parts of the country.

Allen and Burgess (2010) provide a thorough account of the state of the literature on the effectiveness of quasi-market schooling in England. They identify four necessary conditions for a quasi-market system to be effective:

- “Parents must value and be able to correctly identify educational success as a school characteristic.
- “Parental choice must be meaningful and capable of affecting the allocation of pupils to schools.
- “Schools must find it beneficial to be popular and to grow.
- “The best way for schools to be popular must be to raise the quality of teaching and learning, rather than engage in other activities (such as cream-skimming easy to teach children).” (Allen and Burgess, 2010; pp. 5—7).

This report, focussing on the first of these necessary conditions, looks for variation in the structure of London’s, and other cities’, markets by examining geographical variation in the willingness of families, and particularly disadvantaged families, to seek out high-performing and effective schools, even if it means travelling further. Related to the authors’ second and fourth conditions, we also consider how parents use admissions systems, and whether the design of the admissions system may play a role in the London advantage.

We do not directly address the third condition. It is important to note that constraints in capacity, especially in urban school districts where school estates often have little spare room for expansion, may curtail the ability of popular schools to grow in response to demand. Analysis of differences in actual and potential capacity at demanded schools is outside the scope of this report.

The report pays special attention to variation between different groups of pupils by disadvantage (represented by Pupil Premium eligibility), ethnicity and ability. Thus the analyses are open to the possibility that some or all findings may be attributable to the unique demographic diversity of the Capital.

2.2 School admissions and equality of access

It has been observed that the intakes of autonomous[4] schools, and Church schools in particular, tend to be more socially-advantaged, higher-ability and less ethnically-diverse, than the composition of such schools’ local neighbourhoods (Andrews and Johnes, 2016; Cantle and Kaufmann, 2016; iCoCo *et al.*, 2017).

[4] In this report, ‘autonomous’ schools are defined as those schools that act as their own admissions authorities. This includes academies, free schools, foundation schools and voluntary aided schools.

However, it is not easy to disentangle the causes of the apparent stratification in autonomous schools, including the majority of Church schools. It is possible that the patterns of uneven sorting along socio-economic and ethnic lines merely reflect the diverse preferences for Church schools of different groups. Conversely, it is possible that for some reason disadvantaged and minority ethnic groups are less able to access places at Church and other autonomous schools.

Allen and West (2011) suggest that the socio-economic profile of Christian families in England may explain much of the stratification, since higher-socio-economic groups are more likely to identify as Christian. However, they also find that lower-income families identifying as Christian are less likely to attend a Church school than more affluent families.

Even if admissions arrangements play a role, this may not be to do with deliberate selection. It may be that lower-income groups are less able to access popular schools because they are less able to afford houses within the catchment zones of popular schools, where house prices have been driven up by demand for the schools themselves. This effect, dubbed ‘selection by mortgage’, potentially affects admissions to any popular school whose oversubscription criteria include a distance criterion.

The report uses a statistical method that controls for both variation in preferences, and home location, to isolate and examine families’ different chances of successfully accessing Church schools and other autonomous schools, for children with different characteristics.

3 The Data

To gain insights into choices and selection, the project capitalises on the newly-available parental preferences data for England. The full dataset consists of records for all children who were in year 6 in the academic year 2013/14, and applying to an English state-maintained secondary school for entry in September 2014. For each child we have a record listing up to six schools that were ranked by the child's parents as part of the Local Authority's co-ordinated allocations process. We also have the child's Key Stage 2 (KS2) results and year 6 census record, and the linked year 7 census record for 2014/15, where available. The school identifiers also allow linking to school performance tables (containing school-level GCSE performance measures), the school-level census data (for demographic information), historical Ofsted data, and to Edubase, the public database of schools' information.

The preferences dataset identifies the schools listed by each parent, and also identifies the school that was offered to the parent at the conclusion of the admissions process. From this we can determine the family's chosen school, whether or not the child gained admission to that school, and the rank of the school that the child *did* gain admission to. The linked datasets provide the child's location, prior attainment, ethnicity, Pupil Premium status and other characteristics, as well as identifying the secondary school the child actually enrolled in, if the child stayed in the state sector.

3.1 The secondary school admissions process

Each year, parents in England who wish to send their children to state-maintained secondary schools are invited to rank between three and six schools in order of preference. Parents are encouraged to use all of their options, but many parents submit only one or two preferences. Local authorities receive the ranked lists and inform schools of the applications they have received[5].

Whenever the number of applications received by a school exceeds its Published Admissions Number (PAN), the quota set by the school for admissions, the school is asked to submit to the local authority a ranking of applicants according to the school's published admissions policy. For community schools the local authority creates the ranking. Own-admissions authority schools, such as academies and voluntary-aided schools, use information they have gathered through the admissions process and supplementary information forms to rank applications themselves.

[5] Note that, since the Admissions Code of 2007, schools are not informed of the parents' rankings of the school, so that they cannot prioritise, say, first preferences above second or third preferences.

Local authorities then use parents' ranked lists of schools, and schools' ranked lists of children, to allocate school places fairly and ensure that every child is allocated to a school. This part of the process is not as straightforward as it sounds, and in fact the efficient solution to the problem of reconciling all of these preferences and priorities draws upon a substantial body of theory in the fields of economics and computer science[6].

Although it is not necessary to go into the details of this theory, it is sufficient to recognise that there may be more than one plausible allocation for a given set of preferences and priorities, and that the details of the *allocation mechanism* affect not only the outcomes of the admissions process, but also, potentially, the incentives that parents (and admissions authorities) face in submitting their lists of preferences (priorities). In short, each local authority uses an algorithm[7] to reconcile the competing rankings of parents and schools to compute an allocation that satisfies some criteria of fairness and maximises satisfaction. After parents have been informed of their child's allocation, they may appeal the decision or request to be added to their preferred school's waiting list.

3.2 The impact on analyses of strategic choices

When asked to provide a ranking of secondary schools, parents are expected to order schools according to their underlying preferences for those schools. The allocation mechanism then uses both parents' rankings of schools, and schools' rankings of pupils (according to their admissions policies) to compute an allocation of pupils to schools. However, the assumption that parents rank all schools according to their true preferences is a strong assumption, and there are several plausible ways in which the assumption may be violated.

First, the implementation of the admissions system may be such that parents are disincentivised from listing their true preferences (in the sense of increasing the risk of receiving a less-preferred allocation), and second, there may be widespread misconceptions about the admissions process that mean that even when it is in parents' best interests to rank schools straightforwardly, they may not be aware of this.

The strategic incentives that parents face depend to some extent on the details of the admissions system and particularly the allocation mechanism. Although there is no recent survey of all allocation mechanisms used by local authorities, the majority of local authorities apparently use a variant of the Gale-Shapley Deferred Acceptance (DA) algorithm to compute an allocation (Gale and Shapley, 1962). Under ideal conditions, this algorithm has a 'strategy-proofness' property, which means that parents cannot improve

---

[6] The branch of economics known as mechanism design uses game theory to analyse the properties of allocation mechanisms.

[7] The specific algorithm used depends upon the local authority.

their children's allocation by ranking schools strategically, for example by taking into account acceptance probabilities when including or ordering schools. In other words, parents should *not* take into account admissions policies, and should just consider their rankings *as if* they were guaranteed entry to any school. Anecdotally, this appears to run counter to some parents' understanding of the admissions process.

However, the DA algorithm as implemented in England may fail to achieve this 'strategy-proofness' property, because local authorities limit the length of parents' preference lists. Under these conditions, where slots are rationed, parents may need to select schools strategically to avoid allocation to a school they did not express a preference for. If children fail to gain admission to any of their stated preferences, they will typically be allocated by the local authority to the nearest school with spare capacity, which may not even be near to the child's home. The potential severity of this problem is inversely proportional to the maximum length of preference lists.

So, the issues affecting parents' incentives to list schools according to their true preferences, relate to both the design of admissions systems, and confusion about the details of the process. The possibility that parents are, to some extent, second-guessing chances of admission when ranking schools, must be taken account of when interpreting the results of discrete choice analyses. It must also be taken account of when using admissions statistics to gauge the level of parental satisfaction with the choice process.

3.3 Choice of locations

The full record for 2013/14 contains about half a million records. Due to the complexity of the statistical models to be estimated, it is infeasible to use the entire dataset. Instead we select a sample from Greater London, and choose three other locations for comparison. Greater Manchester and the metropolitan area around Birmingham were selected, as the largest metropolitan areas in England apart from London. Both cities have dense school markets with diverse secondary school provision, and socio-economically and ethnically diverse populations.

Providing a less metropolitan comparison is more difficult, as the sample sizes of socio-economic and ethnic minority groups tend to be very small outside large cities. However, it was felt to be important to include such a comparison to evaluate the extent to which the 'London effect' might be more generally a 'metropolitan effect'.

The area of West Yorkshire and East Lancashire straddling the M62 corridor in between Manchester and Leeds, was selected as it contains a number of small towns with socio-economically and ethnically diverse populations close to, yet geographically and economically distinct from, the neighbouring cities. The area overlaps with the North Eastern boroughs of Greater Manchester: Bury; Rochdale; and Oldham. Three towns in the area, Bradford, Burnley and Oldham, suffered riots in 2001 that brought attention to ethnic divisions in the area. At the time, ethnically-segregated secondary schools were cited as a possible contributing factor to the problem (Cantle, 2001; Ouseley, 2001). The sample includes two towns recently designated ‘education opportunity areas’ – Bradford and Oldham. This sample is named ‘Pennines’ because the sample area falls within the Pennines upland region of Northern England.

The London sample comprises a random sample of children living in one of the 32 London boroughs and applying for a secondary school place in 2013/14. The initial sample size was 25,000, from a total cohort of about 78,000. The Manchester sample comprises all children living in the local authorities comprising Greater Manchester: Bolton; Bury; Manchester; Oldham; Rochdale; Salford; Stockport; Tameside; Trafford; and Wigan. The Birmingham sample encompasses children living in: Birmingham; Coventry; Dudley; Sandwell; Solihull; Walsall; and Wolverhampton. The Pennines sample encompasses the local authorities of Bradford, Kirklees, Calderdale, Bury, Rochdale, Oldham, Blackburn with Darwen and part of Lancashire to the East of Blackburn. Note that the schools in each sample may fall outside these areas, since children within the areas are permitted to apply for schools in other local authorities.

3.4 Demographics

Table 1 shows the ethnic composition of each of the samples, using four broad ethnic categories (see section 5.4. for a full description). All of the samples have larger minority ethnic populations than the rest of England. However, London stands out as having a far more ethnically diverse population of children than the other areas; it is the only location in which the white British ethnic group is not a majority, and indeed the relative proportions of the four groups in London are almost equal.

The South Asian category has large proportions in all four areas, with the Birmingham and Pennines samples having higher proportions of South Asian pupils than London. In contrast, the population of children belonging to the black category appears to be more geographically concentrated in London and Birmingham.

Table 1: Ethnic mix (%) by location

%	White	Black	S. Asian	Other
London	30	26	23	21
Birmingham	51	12	29	7
Manchester	70	6	17	6
Pennines	68	2	26	4
Rest of England	84	3	6	6

The distribution of children eligible for Pupil Premium (Table 2) is much less geographically varied. The three metropolitan samples all show a rough 60:40 split, whereas the Pennines sample is closer to the rest of England at about 70:30.

Table 2: Pupil Premium proportions (%) by location

%	Not Pupil Premium	Pupil Premium
London	60	40
Birmingham	59	41
Manchester	64	36
Pennines	68	32
Rest of England	74	26

3.5 Characteristics of schools

Each of the four sampled areas contains a mix of autonomous and community schools, a small proportion of grammar schools, and some Church schools. In the context of the report, autonomous schools are those governance types that set and administer their own admissions policies, otherwise known as own-admissions authority schools. This includes academies (of all types), free schools, foundation schools and voluntary aided schools, but not voluntary controlled schools or community schools. Throughout the text, "Church schools" refers only to the two most common denominations – Church of England (C of E) and Roman Catholic (RC). The proportion of other denominations is very small. In most cases Church schools and grammar schools are also autonomous schools.

Table 3 shows that the proportion of autonomous schools is highest in Birmingham, and smaller in the North West. There is a greater proportion of grammar schools in the Birmingham area, and the smallest proportion of grammar schools is in the Pennines. The largest proportion of Church schools is in Manchester.

Table 3: % of schools of different types by location

%	Autonomous	Grammar	Church
London	80	4.1	24
Birmingham	84	5.9	16
Manchester	69	4.3	29
Pennines	70	3.4	25
Rest of England	81	5.5	14

NB: types are not mutually exclusive and so do not sum to 100 in each row. For example the majority of grammar schools and Church schools are also their own admissions authority.

Figure 1 shows clearly that the Ofsted performance of schools in London surpasses the other areas. Birmingham has a slightly larger proportion of Outstanding schools than Manchester or the Pennines, but all three areas have a similar proportion of schools that are either ‘Requires Improvement’ or ‘Special Measures’, which in all three cases is higher than in the rest of England.

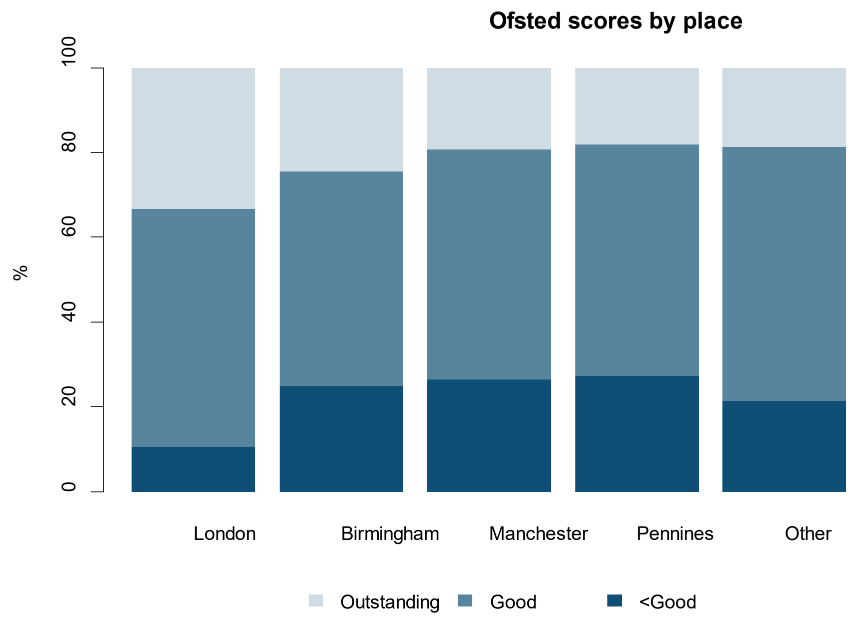

Figure 1: Proportions of Ofsted scores by location

4 Descriptive Statistics

4.1 Proportion admitted to most-preferred school

The headline figure of the proportion of parents receiving their first preference school has, until now, been one of the few pieces of information about each local authority's admissions process that has been made public. The proportion is always high – nationally almost 90% of parents receive their first-choice school.

The national preferences data allows us to disaggregate these headline statistics on 'parental satisfaction' by demographic group. Table 4 shows the proportion of parents receiving their top-ranked school by ethnicity and by Pupil Premium status for each location.

Overall parents are much less likely to obtain their preferred allocation in London (74%) than the national average (88%). In Birmingham also parents are less likely to achieve their preferred school (80%), whereas Manchester and Pennines appear closer to the rest of England in terms of preference rate.

The difference in proportions for Pupil Premium vs. non-Pupil Premium families is quite small or non-existent. In contrast, the difference between white and other ethnic groups is striking in every location. For example, in London 84% of white families achieve their top preference, whereas only 68% and 71% of black and South Asian families, respectively, achieve theirs.

Table 4: % of families achieving most preferred school

	London	Birmingham	Manchester	Pennines	R. of England	England
White	84	89	93	93	94	93
Black	68	72	77	85	80	73
S. Asian	71	68	80	76	80	75
Other	72	78	82	85	86	80
not PP	75	80	89	89	92	89
PP	73	80	88	86	92	86
All	74	80	89	88	92	88

Although striking, these statistics do not necessarily imply that parents in London and Birmingham, and parents of minority ethnicities are less satisfied with their allocation than other parents. It is possible that they are simply more likely to express a risky preference for their first preference, and place a less ambitious 'insurance' school lower down in their rankings, whereas white parents and parents living in other neighbourhoods might be more likely to (erroneously) believe they should name their first preference strategically.

4.2 Number of schools ranked

We can gain some insight into geographical and demographic variations in the operation of school markets by examining the way that parents use the admissions system. Although each local authority maintains its own admissions system, the admissions systems share similar features. The 32 London Boroughs collaborate within the Pan-London Admissions authority, to give parents in Greater London a unified admissions process. Within the Pan-London Admissions system parents may rank up to six schools. Other local authorities outside London must allow parents to rank at least three schools. In practice local authorities allow parents to rank between three and six schools.

Table 5: % of families ranking each number of schools by location

# ranked	London	Birmingham	Manchester	Pennines	R. of England
1	11	25	35	25	42
2-3	37	45	55	65	49
4-6	51	29	10	11	9

Table 5 provides the proportions in each location ranking only one, two to three, or more than three schools. In London parents are far less likely to rank only one school, and far more likely to rank 4-6 schools, than in any other area. More than half of parents rank 4-6 schools, compared to less than one in ten in the rest of England. Birmingham's parents are also more likely to rank more schools than the rest of England. However, in Manchester and the Pennines the proportion ranking 4-6 schools is only slightly higher than the rest of England, although the number ranking only one school is lower.

There are a few possible explanations for these patterns. First, variation in admissions systems may account for much of the variation in parents ranking 4-6 schools. Parents in London can all rank up to six schools, but in Birmingham, Manchester and the rest of England, admissions systems are much more fragmented. Second, the number of viable options for a parent may be expected to depend on the density of the local school market, which in turn will depend on the local population density. Rural areas will have far fewer schools within a certain radius, and this may mean that parents only have one or two viable choices. What is more, parents may be relatively sure of admission to schools in rural areas due to lack of demand from other parents.

Relative admission probabilities may play a greater role in variation in urban areas also. One conjecture is that London seems to have a smaller proportion of schools with formal *geographical priority areas* (aka catchment areas) than in other cities. Without a recent survey of admissions arrangements this is impossible to verify, but if true, it would partly explain the proportions of parents expressing more than one preference, since in this context most families would not be reasonably confident of admission to a local school, and will therefore be more likely to express multiple "insurance" choices.

These considerations mean that it is difficult to draw conclusions from looking at geographical variation in use of the admissions system. We can, however, look at variation in ranking behaviour for different demographic groups, since we might not expect admissions systems to vary greatly within a given location for different socio-economic and ethnic groups.

Table 6 presents proportions ranking different numbers of schools by location, ethnicity and Pupil Premium (PP) eligibility. Within each location, Pupil Premium children's families are slightly more likely to rank only one school, and less likely to rank 4-6 schools. This may suggest less engagement with the admissions process. However, there are other possible explanations. For example, it may be that, due to where they live, or admissions arrangements, Pupil Premium children's families have fewer acceptable options, or are more likely to choose schools that they are relatively confident of admission, such as local schools with spare capacity.

Table 6: % of families ranking each number of schools by ethnicity and Pupil Premium eligibility

# ranked	White	Black	S. Asian	Other		not PP	PP
London							
1	16	8	9	11		10	13
2-3	45	31	35	37		36	39
4-6	39	61	56	52		54	48
Birmingham							
1	32	19	17	25		23	30
2-3	51	43	37	44		47	43
4-6	18	37	46	30		30	28
Manchester							
1	38	25	28	34		33	40
2-3	54	60	57	52		56	53
4-6	8	14	15	14		11	7
Pennines							
1	27	26	16	29		24	27
2-3	67	62	59	61		67	61
4-6	5	12	25	10		10	12

The patterns of ethnic variation in ranking behaviour are more striking. Black and South Asian ethnic groups in particular appear to be more likely to rank more schools, compared to white families. Again, although there are competing explanations for this, a plausible explanation is that these minority ethnic groups are more engaged with school choice than white families. However, geographical variation in admissions systems, the density of local secondary school provision, and relative prospects of admission into local schools are all plausible explanatory factors.

5 Modelling school choices

The aim of the second method is to model parental decision-making when faced with trade-offs between convenience (for example, avoiding large home-school commutes), quality of teaching and learning, peer groups and other considerations.

Tabulating the proportions of parents who choose different schools does not account for the uneven spatial distribution of pupils and schools. Using a statistical model, we can account for these spatial irregularities and uncover patterns in the choices that parents make, given the options that they face.

The main focus of the current analysis will be to produce estimates of parents' Willingness to Travel for improvements in academic performance, and compare these estimates across demographic groups and across locations. Thus we can directly quantify any geographic or demographic variation in parental engagement with school choice.

Previous work to model parental school preferences is limited, because of the rarity of choice data. In the UK Burgess et al. (2014) used the Millenium Cohort Study to conduct a discrete choice analysis. In the US, Hastings et al. (2006), and Abdulkadiroğlu et al. (2015) have used choice data to study the determinants of preferences. This study uses a similar methodology to these previous studies. However, the data at our disposal is of a scale and quality that has not been used before in the UK or the US, and consequently much richer models may be estimated.

5.1 The discrete choice model

A discrete choice model shares the same basic structure as a regression model. There is a dependent variable, in this case a binary variable indicating whether a pupil chose a particular school, and a set of independent (predictor) variables whose influence upon the dependent variable is being modelled. Since the dependent variable is a binary, and not a continuous variable, an extra stage in the model is needed to translate the continuous linear predictions of the regression model into statements about the probability of the dependent variable taking a value of one, which corresponds to the family choosing the school.

The convention in discrete choice modelling is to conceptualise the regression model as predicting a latent (unobserved) variable called *utility*. Each family is assumed to have a separate utility value for each school, which is known to the family but unobserved by the researcher. The family simply chooses the school which gives it the greatest amount of utility. A family's utility for a school is a function of the predictor variables – observed variables relating to the school's characteristics, the family's/child's characteristics, and interactions of the two – plus a random idiosyncratic element (akin to the error term in a regression). Such models are thus known as *random utility models*. The model used here is described more fully in Appendix 4.

5.2 Convenience-Performance trade-offs in school choice

The estimated coefficients of a logit discrete choice model can be interpreted as log odds ratios, relating odds of a pupil choosing a given school when a particular variable in the model is at a particular value, to the odds of choosing a school after a unit change in that variable. However, this interpretation is not intuitively easy to use. Additionally, as the number of demographic groups for whom the coefficients are separately estimated increases, the interpretation of interactions between variables and pupil group identifiers becomes increasingly complicated.

We require a straightforward representation that relates the estimated choice model to parental decision-making, such that the relative strength of demand for academic performance is revealed. Typically in discrete choice contexts where money is involved, coefficients can be converted into monetary values, known as *Willingness to Pay*. In the context of school choice we can use distance, rather than money, as a yardstick against which to measure preferences for test scores. The idea is that parents may face a trade-off between seeking a school with high academic performance, and the convenience of settling for a closer school. Note that for a given family no trade-off between distance and performance may be necessary – the best school may also be the closest – but the model allows the estimation of trade-offs that parents *would* make, if they had to.

A parent's *Willingness to Travel* (WTT) can be interpreted as an approximation to the additional distance that a parent would be "willing" (willingness here meant in an economic sense as encompassing willingness and ability) to travel for a 10 point improvement in test scores (say, from 70% achieving 5+ A*-C at GCSE to 80%). It is described more fully in Annex 4.

5.3 Non-randomness of home location

In interpreting the results of the choice model, and especially in interpreting estimated *Willingnesses to Travel*, the non-random assignment of pupils to home locations must be considered. The locations of family homes are not completely independent of the quality of the local school market, since families choose where to live partly on the basis of the quality of the local school market. We do not model this choice process. It is possible that part of a family's overall Willingness to Travel for academic quality has been subsumed within previous residential moves, for example by compromising between the school commute and work commutes in choosing where to live. This is, of course, mediated by house prices, which may go up or down in response to demand for nearby schools.

If there are systematic variations in the ways that different demographic groups choose where to live, whether because of differences in ability to pay or for any other reason, the estimated *Willingness to Travel* may reflect these differences, rather than underlying attitudes to schooling. For example, a demographic group may appear to be less willing to travel, whereas in fact they are more likely to choose to live in an area that is already near to preferred schools. What this means is that we can only interpret *Willingness to Travel* relative to the residential distribution of demographic groups.

5.4 Specification of the model

The main aim of the model is to quantify the parental valuation of the academic performance of the school, and the trade-off that parents may have to make between academics and convenience (proximity). Of particular importance is the extent to which this trade-off is evaluated differently by families from different socio-economic and ethnic groups, and taking into account different levels of prior attainment.

To this end, the pupil utility model must take account of academic performance as measured by test scores, as we wish to evaluate the extent to which families prefer higher-performing schools. It must also take account of proximity, as parents have obvious incentives to prefer closer schools over schools further away, all things being equal. The NPD data contains the postcodes of both schools and pupils as well as the Ordnance Survey co-ordinates of postcodes, measured at the geographical centre, or centroid, of the postcode area. The OS co-ordinates are used to calculate the linear distance between each pupil's postcode and each school in kilometres.

We include both distance and the natural logarithm of distance in the model. Taking the log of a variable is a common variable transformation in econometric analysis. Whereas the coefficient of an untransformed variable is interpreted as a measure of the decision-maker's consideration of a unit change in that variable, the coefficient of a logged variable can be interpreted as the decision-maker's sensitivity to a proportional change in that variable. For example, a model including only a linear term for distance would impose the assumption that a parent cares as much about a difference between a school 1 km away and a school 2 km away, as she does about the difference between a school 21 km away and a school 22 km away. The coefficient on the logged variable, in contrast, assumes that a parent cares only about the proportional change in distance, so that a 10% increase in distance at 1 km (to 1.1 km) matters as much as a 10% increase in distance at 20 km (to 22 km). Including both of these terms in the model allows for both types of consideration, or a mixture of the two.

With regards to measuring academic performance, there are several possible alternatives. During the period in which the 2013 intake were deciding upon secondary schools, the most high profile measure of test score success was the proportion of pupils achieving five or more A*-C grades at GCSE (AC5). Other measures include GCSE point scores and value-added measures (Best 8 value-added). However, the value-added framework was in a transitional state in 2013/14, and the 'Best 8' VA measure was not well-publicised, therefore parents may not have been sufficiently aware of the value-added measure to take it into account in decision-making. Ofsted scores are another high-profile measure of school effectiveness. However, Ofsted scores have a broader focus than just academic performance, and additionally only have four levels, so are not finely differentiated enough to provide the main measure of academic success. However, Ofsted scores are included in the model to explain residual variation in popularity after controlling for test scores.

The AC5 variable originally takes values in between 0 and 100, and is scaled to take values between 0 and 10, so that a unit change in the variable corresponds to a change of 10 points in the proportion of pupils achieving five or more GCSEs at A*-C.

Since a 'raw' measure of exam success has been used as the main measure of academic performance, it is important to consider the ways in which scores are correlated with the characteristics of the intake. Some secondary schools predominantly take in pupils who have good attainment prior to starting at secondary school, either because of entry criteria, or demographic variation in their catchment areas. Other schools' intakes have lower prior attainment. One reason for taking this variation into account is to allow for the possibility that what parents really care about is 'value-added' – that is, the academic performance of pupils relative to their starting points. It is possible that parents may consider this to be a more reliable indicator of underlying teaching quality. It is also important to consider the converse, that what parents really care about is the 'quality' of peer groups.

We can use the average Key Stage 2 points score (KS2 APS) of a secondary school's intake as a measure of the academic characteristics of the peer group. The average KS2 points score summarises the prior attainment of the cohort before they joined the secondary school. This is highly correlated with the same cohort's AC5, and can be used to decompose the relative contributions to AC5 of 'value-added' and the 'peer group'. To achieve this, we regress AC5 on a flexible function of KS2 APS, and decompose AC5 into predicted values (representing the effect of the prior attainment of the intake) and residuals (representing a crude 'value-added' measure).

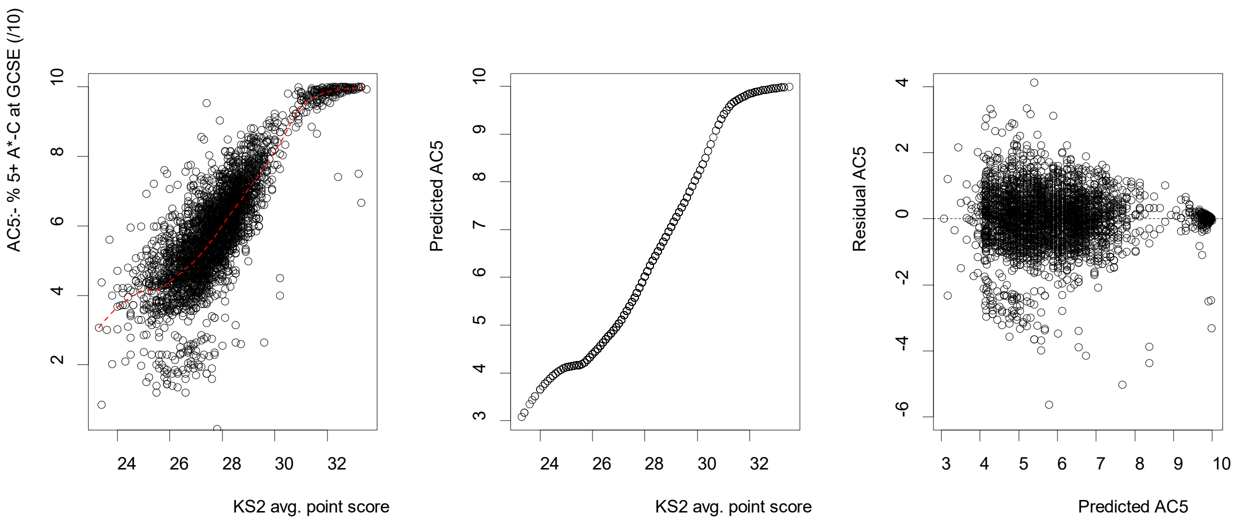

Figure 2: Decomposition of AC5 into predicted value and residual on KS2 avg. point score

The resulting variables – AC5_predicted and AC5_residual, which are completely uncorrelated – can be included in the discrete choice model at the same time to assess the relative importance attached by families to academic performance due to 'peers' and academic performance due to 'value-added'. Following the decomposition, both AC5_predicted and AC5_residual are on the same scale as AC5, as shown in Figure 2, so that it is possible to compare the coefficients on the measures directly.

There are therefore two models of interest: one which includes distance, logged distance and AC5; and the other which includes distance, logged distance, AC5_predicted and AC5_residual.

In each of these models the coefficients for these main variables are allowed to vary by the pupils' Pupil Premium status, ethnic group and prior attainment. Pupil Premium status is defined, for the purposes of this study, as an indicator of whether the child has been eligible for Free School Meals at any point in the past six years, which may be slightly different than the formal definition of Pupil Premium. This indicator is taken from the NPD. The detailed ethnicity data in the NPD was grouped into four categories: white British (including white Irish but excluding 'white other'); black (including black Caribbean and black African, and mixtures of these); South Asian (including Indian, Pakistani and Bangladeshi, and mixtures of these); and other (including Chinese and 'white other'). This coarse ethnic grouping ensures large enough sample sizes in each of the groups.

To group pupils on prior attainment, each pupil's KS2 average point score was binned into three categories, corresponding to the lowest 30% of scores, the middle 40%, and the highest 30%.

For each ethnic group, for each KS2 group, and for Pupil Premium pupils, a separate coefficient was estimated for distance, logged distance and AC5 (or AC5_predicted and AC5_residual).

Finally, variables indicating the characteristics of the school are included to 'mop up' some of the remaining variation in school popularity, not explained by test scores. Ofsted scores are grouped into (1: 'Outstanding', 2: 'Good' and 3: 'less than Good'). Since only small numbers of schools fall into the two categories 'Requires Improvement' and 'Special Measures' they have been grouped. In addition, Ofsted provides a measure of the deprivation of each school's neighbourhood, based on Indices of Multiple Deprivation (IMD), an area-based measure. Three IMD groups ('High', 'Medium', and 'Low' – where 'Low' is the least deprived). A separate coefficient is estimated for each combination of Ofsted score and deprivation group, allowing for interactions between Ofsted success and local deprivation in explaining school popularity. Indicator variables are also allowed for Church of England schools and Roman Catholic schools, to quantify the relative popularity of these denominations. The main purpose of these fixed effects is to control for variation on these characteristics; they are not the focus of analysis.

5.5 Choice model results

The estimated coefficients from the discrete choice models are presented in Table 8 and Table 9 in Appendix 1. The estimated coefficients on AC5, distance and logged distance have been used to calculate *Willingness to Travel* (WTT), representing the additional distance that a parent would be willing to travel to school, according to the model, for a 10 percentage point improvement in test scores.

Figure 3 presents the Willingness to Travel for London, for families with children with medium prior KS2 attainment. In this figure and in all subsequent figures, error bars extend two standard errors on either side of the estimate, and represent an approximate. 95% confidence interval for the estimate.

A non-Pupil Premium white family would be willing to travel about a quarter of a kilometre further for a school with AC5 10 points higher. However, according to the model a Pupil Premium white family would not be willing to trade any convenience for test scores at all.

There is substantial variation by ethnicity; South Asian and black families without Pupil Premium would travel about 2.5 times as far as white families for a 10 point improvement in test scores. As discussed in Section 5.3, this difference may be mediated by variation in the residential choice process. It is possible, for example, that white families place more emphasis on schools when choosing where to live, so that they do not need to travel as far at admissions time. It is also possible that mean incomes differ for the ethnic groups, so that some groups are more able to pay higher prices to live in the catchment areas of desired schools.

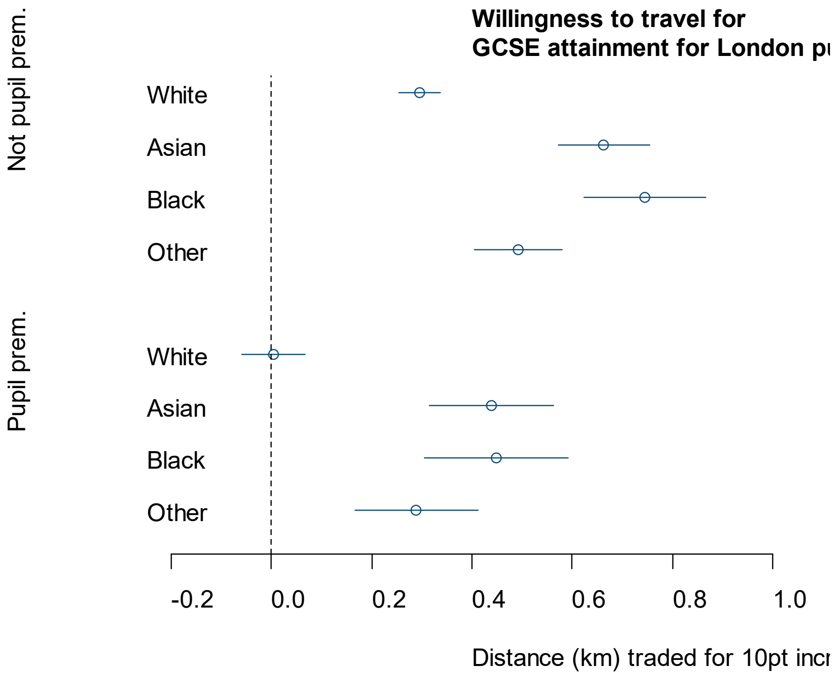

Figure 3: Willingness to Travel at 2.5km. London only, medium KS2.

Figure 11 in Annex 2 shows the WTT for all locations, broken down by prior KS2 attainment of the child. The middle panel of Figure 11 reproduces the results for London, which are not obviously exceptional in the context of the results for the other cities. Interestingly, there does appear to be a difference in WTT between London and Manchester, on the one hand, and Birmingham and the Pennines, on the other. In general, the model predicts that families of high-KS2 children would be willing to travel further in all cities. In London this interaction is especially pronounced.

Model B decomposes AC5 into two uncorrelated constituent elements: the part predicted entirely by the average prior attainment of the cohort; and the residual that is not predicted by the intake. The residual can be thought of as a crude value-added measure.

The WTT from model A are compared to the two sets of WTT from model B in Figure 7 to Figure 10. The results for London in Figure 7 are broadly representative of those for the other locations. The main pattern is that in the middle panel (medium KS2 child) the WTT based on AC5_predicted is not significantly different from the WTT based on AC5_residual, and neither are significantly different from the WTT based on AC5. Since all three measures are on the same scale these WTT are directly comparable. This suggests that these parents pay attention to test scores, but do not differentiate between schools that have high performance because of their intake, and schools that have high performance because of value-added (due, for example, to effective teaching practices).

However, the rightmost panel, for high KS2 children, shows that the differentiation of WTT by the prior attainment of the child is much more muted when considering AC5_residual, than either of the other two measures. This may be because high-KS2 children's parents seek peers with similar prior attainment. Conversely, it may be explained by strategic ranking behaviour. Schools with a high AC5_predicted must have a high-prior-attainment intake, and probably many of these are academically selective. Parents of pupils with lower KS2 attainment may avoid applying for these schools because they perceive their child will not be admitted.

The estimated WTT for the Pennines sample (Figure 10) are qualitatively different than the other locations, in that it is the only location in which minority ethnic families do not have greater demand for test scores, relative to proximity, than white families. WTT for all ethnic groups are lower than in Manchester and London, and especially so for South Asian families.

5.6 Discussion

Although there is some evidence of subtle variations in choice behaviour for different locations, these variations do not appear to be important compared to the variation in choices between socio-economic and ethnic groups.

In broad terms, Pupil Premium children's families seemed less likely to travel further to seek out academic performance than non-Pupil Premium children's families, and minority ethnic families, particularly black and South Asian ethnicities, seemed more likely to travel further seek out academic performance than white families. The main exception to this pattern is in the Pennines, where South Asian families did not seem to travel further for improvements in test scores than white families.

To what extent do these patterns represent genuine variation in preferences, and to what extent do they reflect strategic choices by some or all groups of parents? The results control for variation in pupils' prior attainment, so admissions rules focussing on aptitude or ability banding (which would be expected to be more prevalent at high-performing schools) should not be an important issue in explaining the different propensity of different groups to travel for performance improvements.

And yet, as Table 4 shows, minority ethnic groups face significantly smaller probabilities of accessing their first choice school. It is impossible to answer this question without quantifying the different chances of access to desired schools that different socio-economic and ethnic groups face, which is the aim of the case-control matched analysis of school admissions selection.

6 Case-control analysis of admissions selection

The aim of the current analysis is to examine the effect of school admissions practices on the chances of different groups being able to access chosen schools. It has been observed that popular schools, and particularly certain types of own-admissions authority schools, tend to have intakes that do not reflect the socio-economic, ethnic or ability composition of their local catchment areas (Andrews and Johnes, 2016; Cantle and Kaufmann, 2016; iCoCo et al. 2017).

Although it has been suggested that school admissions practices may play a role in this apparent sorting it has not been possible, with existing data, to disentangle the effect of admissions arrangements from the choices that parents make between schools. The new preferences dataset enables choices and chances of admission, having made a choice, to be isolated to examine their relative contribution to stratified intakes.

This is because, given we know the eventual offers that were made, information on rankings on one side of the market reveal information about rankings on the other side of the market (Weldon and Titman, 2015). Intuitively, if we observe that a family has named a school as first preference, but the child has not been offered a place at that school, then we know that the school ranked all of the pupils offered a place at the school higher than that child. This allows us to identify a set of 'unadmitted' pupils for each school, who we know were ranked lower than the 'admitted' pupils who were offered a place at the school.

With this information, it is possible to model the probability of acceptance to a school. However, the most important determinant of access to schools is often the home location of the child's family. The ubiquity of catchment areas and distance 'tie-breakers' in admissions policies has been dubbed 'selection by mortgage' since it is believed that admissions policies affect house prices, causing disadvantaged families to be priced out of the catchment areas of popular schools. We wish to study the remaining selection that affects families, after 'selection by mortgage' is accounted for. For this reason we adopt a design that matches nearby pupils to each other – unadmitted matched to admitted – so that the effect of distance to the school does not need to be considered. This 'case-control' design matches 'cases' – unadmitted pupils – to their nearest 'controls' – admitted pupils – allowing many-to-one matches where possible.

A possible problem with this approach is that the 'cut-off' – the furthest distance at which a school will accept pupils – is a sharp divide, and even pupils matched to other pupils very close (on the same street) may fall either side of this cut-off. This may introduce bias, as unadmitted pupils may tend to be slightly further away from the school than admitted pupils. In order to completely remove the possibility of this bias we can consider *only* case-control matches where the admitted pupil(s) are *at least as far* from the school as the unadmitted pupil. Figure 4 represents this case-control matching design. The matching radius sets the maximum distance between matched pairs. Within the radius, the matching is selected to minimise the distance between cases and controls.

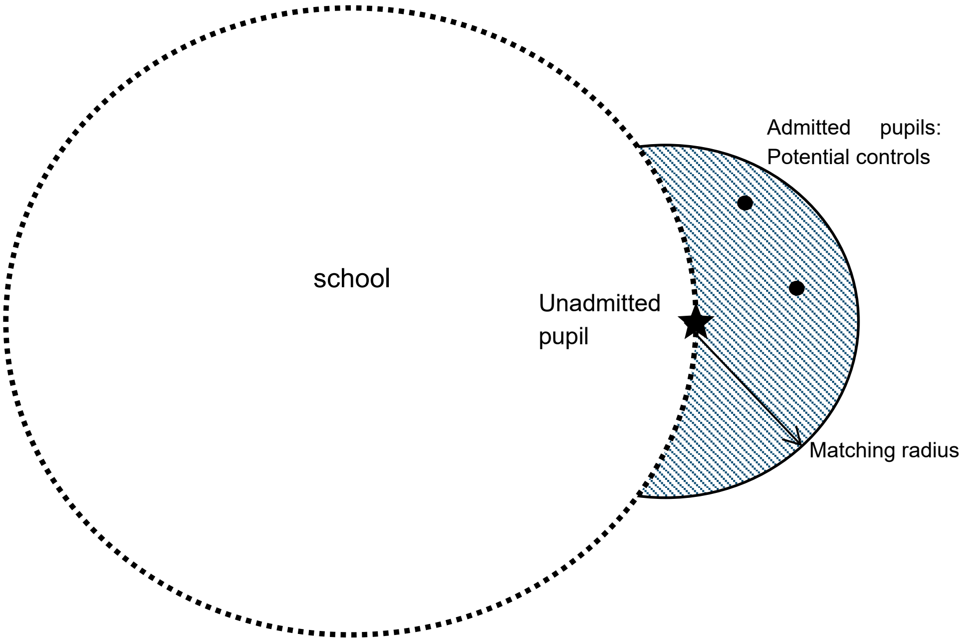

Figure 4: Design of 'case-control' matching of unadmitted to admitted pupils

6.1 The matched case-control model

The model we specify shares a similar structure to the discrete choice model described in the previous section. In fact, it is useful to consider the analysis as analogous to modelling the way that schools "choose" which pupils to admit or reject from a "choice set" of pupils that are geographically close to each other. When making this analogy, it is important to be clear that we are not assuming that schools have agency in this choice process, or consciously select according to demographic characteristics, merely that the model for selection is similar to a model for choice.

Appendix 5 more fully describes the model. The estimated coefficients can be interpreted as the logged odds ratios of rejection for a pupil with a given characteristic, compared to a 'baseline' pupil: white; medium KS2 attainment; not Pupil Premium; no Special Educational Needs. Perhaps a more intuitive way to interpret the coefficients is by imagining that a school has only one seat remaining, and a pupil *A* with a given characteristic, say Pupil Premium, is assessed for that one seat alongside a nearby 'baseline' pupil *B*. We can calculate the probability that pupil A will be ranked lower, and hence fail to be admitted, in this hypothetical scenario. If Pupil Premium status is not related to admissions success we would expect the probability to be 0.5.

6.2 Specification of the model

Schools' admissions policies tend to include various elements. Some of the more common elements are:

- Priority for looked-after children (this is compulsory as of Admissions Code 2007)
- Priority for children for whom there is a medical need to attend
- Priority for children with older siblings already attending the school
- Priority for children living within a geographical priority area (GPA)
- Priority for children attending certain feeder primary schools
- Some schools select a small proportion on aptitude (e.g. musical or languages)
- 11+ selection by grammar schools
- Religiously designated schools may prioritise children demonstrating adherence to a certain faith. Evidence required varies greatly, but may require, for example, validation by a minister of the faith
- Banding on academic ability, whereby a school allocates quotas for each prior attainment band
- After these criteria have been applied, Some measure of distance from the school is the most common criterion used to break ties between children with the same priority.

There is scant information in the data to model these admissions criteria. We do not know about siblings, faith, feeder primary schools, GPAs, or aptitude. What we can do is model the average effect of these admissions criteria on our different family groups. That is, in aggregate, do admissions policies tend to differentially affect Pupil Premium, or non-Pupil Premium pupils, different ethnic groups, or different ability groups?

Looked after children are excluded from the analysis, as schools are legally obliged give them top priority. Similarly, although the law is more ambiguous about 'medical reasons' for acceptance, pupils with Special Educational Needs (SEN) who have a Statement[8] have been excluded as their allocation is usually jointly decided by the parents, schools and local authorities. Non-statemented SEN pupils have not been removed but their status is controlled for in the model.

Grammar schools are also excluded from this analysis, since their method of selection means that families are often aware of whether or not they have met the requirement

---

[8] Most children with SEN do not have a Statement of SEN. Obtaining a Statement of SEN provides recognition of more complex needs and can facilitate access to funding and assistance, including consideration of needs in school admissions. Statements are being replaced by Education, Health and Care (EHC) plans.

before they state their preferences. Each child's KS2 attainment, binned into three categories as in the previous analysis, is used as a proxy for the academic ability of the child. Schools do not have access to this measure as the tests do not take place until after the allocation is finalised. However, some autonomous non-Grammar schools make use of aptitude tests that assess aptitude in a prescribed subject, and so admission into oversubscribed autonomous schools might be expected to be correlated with KS2 scores.

We split schools into three categories: non-denominational autonomous schools that set and administer their own admissions policies (includes foundation schools, academies, voluntary-aided schools and free schools); autonomous Church of England (C of E) and Roman Catholic (RC) Church schools, the majority of which have some faith criteria in their admissions policies; and community schools whose admissions policies are set and administered by the local authority (includes community schools and voluntary-controlled schools). The small number of faith schools with denominations other than C of E and RC are included in the non-denominational category according to their governance status. In 2013, about 80% of secondary schools have admissions autonomy (cf. Table 3). Separate coefficients are estimated for Church, other autonomous and community schools.

6.3 Results of case-control model

The estimated coefficients from the matched logit model are presented in Table 10 and Table 11 in Appendix 3. The first aim of the analysis is to determine whether the relative access of different pupil groups to schools is qualitatively different in London compared to other cities.

The model also estimates separate interactions for each school type, which means that it is also possible to compare the effects of admissions policies at Church schools and autonomous schools to those at community schools, to determine if admissions autonomy leads to less equitable outcomes for different demographic groups.

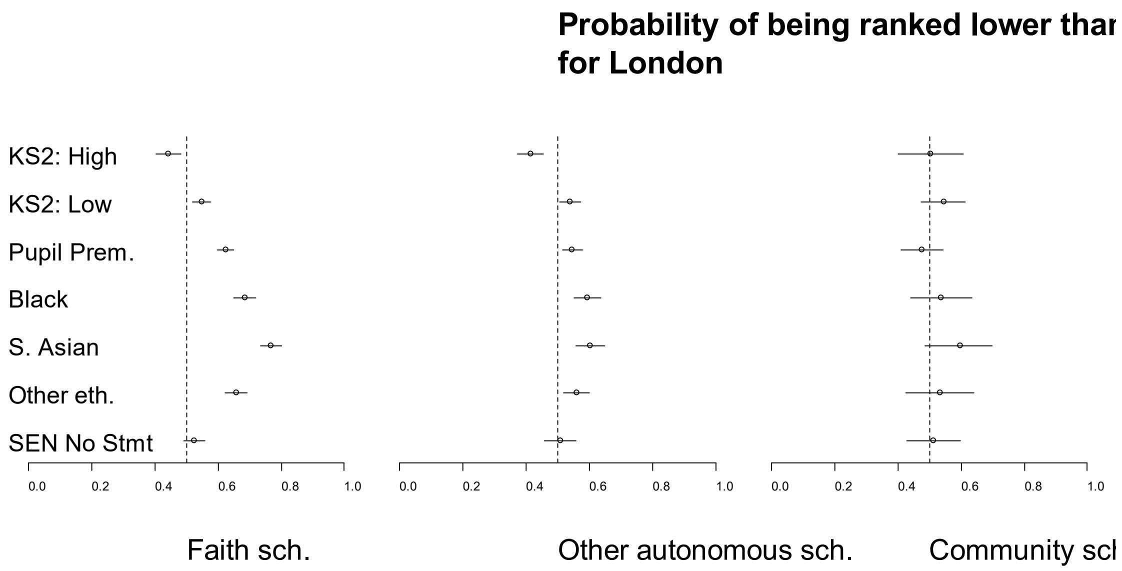

Figure 5: Estimated probabilities of being ranked lower than a white, medium-KS2, non-Pupil Premium child at a similar home location, for admissions in London using a matching radius of 1000m.

Figure 5 presents the probability estimates for London. The probability of being ranked lower than a 'reference' pupil is also the probability that a pupil with a given characteristic will fail to gain the last remaining seat at a school, when assessed for that seat alongside the reference pupil. If the confidence intervals do not cover 0.5 this means that the characteristic has a statistically significant effect on the probability of admission.

As expected, there is some evidence that children with higher KS2 were more likely to be admitted into autonomous schools, including autonomous Church schools. There is also evidence of differential access for both Pupil Premium pupils and minority ethnic pupils, to autonomous schools and Church schools. However, the estimated coefficients for Church schools are much larger.

In a hypothetical comparison between a Pupil Premium child and a non-Pupil Premium child for a Church school place, the probability that the Pupil Premium child would not be admitted is 0.62. If a South Asian child is assessed for a single seat at a Church school alongside a white child, the corresponding probability that the South Asian child would not be admitted is 0.77.

Figure 6 presents the results for Birmingham and Manchester combined. Although the estimates are noisier they are qualitatively similar. The main difference is that there is less evidence of a correlation between prior attainment and probability of admission at autonomous and Church schools. The estimates of selection on Pupil Premium are also smaller, and not significantly different from zero, for other autonomous schools, whilst for Church schools the estimates of selection effects for Pupil Premium and South Asian pupils are slightly larger than London.

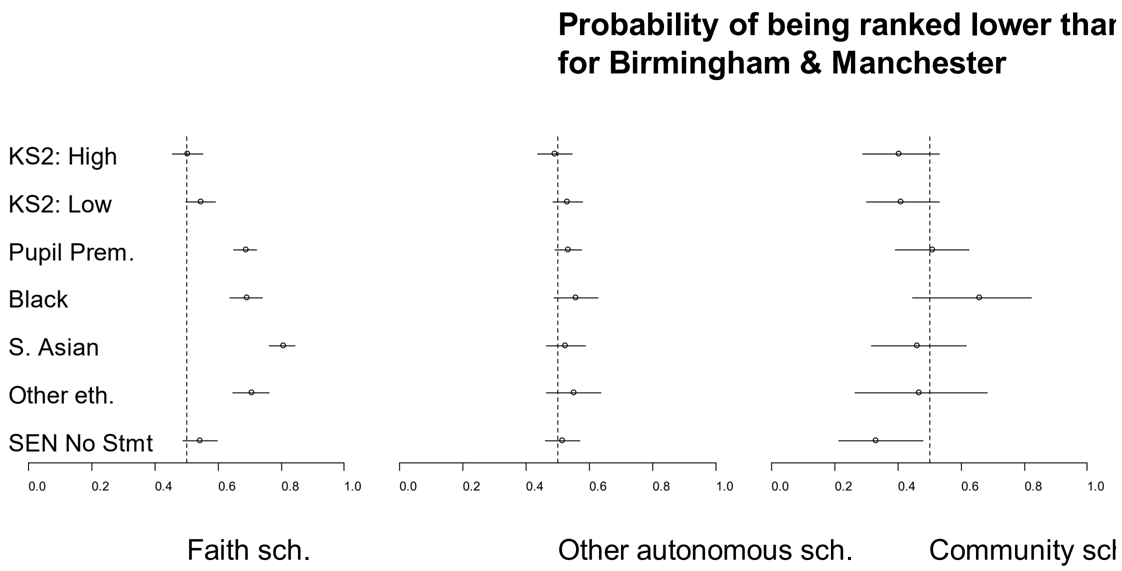

Figure 6: Estimated probabilities of being ranked lower than a white, medium-KS2, non-Pupil Premium child at a similar home location, for admissions in Manchester and Birmingham using a matching radius of 1000m.

As a robustness check, Figure 12 and Figure 13 in the appendix present the same estimates for different matching radii, including 200m, 500m and 1000m. The estimated probabilities become noisier as the radius is reduced due to the decreasing sample size, but the results are not qualitatively affected.

6.4 Discussion

As with the previous analysis, the general picture is of clear differences in the experience of the admissions process, and the application of oversubscription criteria in particular, by ethnic groups that appear to be broadly similar across locations. There appears to be a differential treatment of different demographic groups by some types of school.

These effects appear to be driven primarily by Church schools, although there is more moderate evidence of selection at other autonomous schools in London. Therefore, possible explanations must focus on the admissions practices of Church schools, which are varied but often include the use of supplementary information forms, feeder primary schools, requirements for attendance at religious worship, and countersignatures by members of the clergy.

Given the estimated willingness of minority ethnic groups to travel for academic performance, it is possible that parents of minority ethnicities are more willing to take risks in applying for desired schools for which they have low probability of admission. However, it must be borne in mind that these results control for distance; the difference in admission probabilities cannot be explained by parents choosing schools that are further away. If they are more willing to take risks, it must be in negotiating other oversubscription criteria besides distance.

The low probability of acceptance of South Asian children might be explained by the lower proportion of South Asian families that identify as Christian. More surprising is the lower probability of acceptance of black families who, according to the 2011 census, are more likely to identify as Christian (69%) than white families (64%). However, this does not necessarily imply that the black families that fail to secure admission to their chosen church school are Christian, or that they are practicing members of the relevant denomination.

The results do not imply that Church schools are cream-skimming pupils, nor do they provide any evidence that admissions oversubscription criteria are not being applied in accordance with the Admissions Code. However, the results are striking in that they appear to refute the notion that differences in the composition of Church schools and other types of school can be explained entirely by parents' preferences. As Table 7 shows, black and 'other' ethnicities are, in fact, more likely to choose Church schools than white parents, and South Asian families only slightly less likely.

Table 7: % choosing different types of school by ethnicity and Pupil Premium eligibility

	White	Black	S. Asian	Other		not PP	PP		All
Church	16	27	15	23		17	16		17
Autonomous	64	56	60	59		63	62		63
Community	20	17	25	18		20	21		20

7 Conclusions

7.1 Social mobility and the London effect

The broad picture from the choice analysis is that, although there are some geographical differences in behaviour, it is subtle, and less important than variation between socio-economic, ethnic, and prior-attainment groups in describing the structure of local education markets. London does not seem to ‘stand out from the crowd’ in terms of parental choice behaviour after accounting for demographics, except when it comes to high-KS2 children and their families, and this finding may be to do with strategic choices.

However, London does stand out from the crowd in terms of its demographics. Table 1 shows that the ethnic mix of London’s children is far more diverse than England’s other largest cities. The city’s children are also, thanks to the London effect itself, more likely to have high prior attainment. This means that the aggregate behaviour of parents in London is likely to be more performance-seeking, both in terms of ‘peer groups’ and ‘value-added’. As discussed in the introduction this is a necessary, but not sufficient, condition for the kind of quasi-market competitive mechanism that might drive up standards in the Capital, relative to other cities. So these results are not inconsistent with the ‘market structure’ hypothesis for the London effect. However, it is not clear whether the other conditions for effective quasi-market competition are met in London, particularly the need for sufficient capacity at demanded schools, and financial consequences for falling demand.

The results also indicate persistent differences between ethnic groups that point to different attitudes to educational success. This project is limited in scope and so we have not attempted to offer explanations for these findings, except to caution that differences in preferences between groups are probably due in part to differences in access to schools, and the non-random spatial distribution of families of different backgrounds.

People’s school choices are inextricably linked with their opportunities to access schools. The current admissions system does not incentivise parents to state their preferences entirely transparently, without taking into account probabilities of acceptance. What we observe are possibly strategic choices that combine ‘true’ preferences with assessment of the risks of stating a particular ranking. What is more, our analysis of admissions reveals that different demographic groups face different admission probabilities to the same schools, meaning that the socio-economic and ethnic variation in choice priorities that we observe may also be influenced by strategic choices.

In the light of this, strategic explanations for observed choice behaviour must be considered before other explanations can be entertained. Do Pupil Premium pupils’ families value academic performance less and proximity more than non-Pupil Premium pupils, or are they acting on the belief that that they are less likely to be admitted, and therefore risk wasting an option, if they make choices which prioritise a school's GCSE performance? Do high-KS2 pupils' families value academic performance more, or are we just observing that they are more likely to be confident of acceptance at a school that operates some form of academic oversubscription criterion?

However, it is unlikely that all of the variation in parental preferences can be explained away. The remaining variation in preferences, particularly the ethnic dimension, begs further scrutiny, perhaps employing sociological or ethnographic methods that are outside the scope of this project. Yet it would be worthwhile to go into greater depth in investigating the reasons behind these observed patterns of ethnic variation in preferences, as well as exceptions to these patterns such as in the Pennines, because it is possible that these 'anomalies' contain important insights.

7.2 Admissions practices

The results of the admissions analysis provide evidence that disadvantaged and minority ethnic pupils appear to be less likely to be admitted into own-admissions authority schools, and particularly faith schools.

After controlling for the residential location of pupils, and their prior ability, the remaining differences in rejection probabilities cannot be explained by residential sorting, house prices, or residual (i.e. non-grammar school) aptitude selection.

The estimates of the effect of a school's denomination (in Table 8 and Table 9 in Appendix 1) suggest that Church of England and Roman Catholic schools are not in higher demand than other schools after accounting for test scores. However, faith schools generally perform well, and because of this are popular with minority ethnic families who, it has been shown, tend to value academic performance highly.

Previous research has discussed the socio-economically and ethnically segregated intakes of Church schools (cf. iCoCo et al, 2017) but has not been able to determine whether this segregation was primarily explained by parental preferences or admissions arrangements. These findings imply that the patterns of segregation in Church schools are not explained by preferences, and are, at least in part, due to children failing to gain admission at chosen schools.

7.3 Further work

This project has provided a 'first look' at the newly-compiled parental preferences data, and the insights into the English education market that can be gleaned from it. The author is part of a team embarking on a two-year project, funded by the Nuffield Foundation and led by Professor Ian Walker at Lancaster University, that will be using the data to further explore and flesh out these themes.

We intend to use rich models of school choice in several ways. First, we will evaluate the experience of school choice nationally, to determine the existence of choice 'deserts' and the effect of variations in school choice on the functioning of education markets. Related to this, we will examine the extent of peer preferences – that is, families' consideration of schools' socio-economic and ethnic composition, rather than their performance, in making school choices.

Second, there is potential for models of school choice to be used to improve quasi-experimental designs for the evaluation of education policies. This is because currently school choice behaviour is a black box that potentially biases the assignment of pupils to different schools. Being able to model this part of the assignment process will open up the possibility of accounting for this bias, to enable more accurate evaluations of policies such as the academies programme, or the proposed expansion of grammar schools.

There are therefore several avenues to extend these analyses, and at each stage we will seek advice from, and share insights with the Department.

References

Abdulkadiroğlu, A.; Agarwal, N. & P. Pathak (2015) “The Welfare Effects of Coordinated Assignment: Evidence from the NYC High School Match,” NBER Working Paper No. 21046.

Allen, R. and S. Burgess (2010) “The Future of Competition and Accountability in Education,” report to the 2020 Public Services Trust at the RSA.

Allen, R. & A. West (2011) “Why do faith secondary schools have advantaged intakes?: the relative importance of neighbourhood characteristics, social background and religious identification amongst parents.” *British Educational Research Journal*, 37 (4). pp. 691-712.

Andrews J and Johnes R. (2016) “Faith schools, pupil performance and social selection,” Education Policy Institute report.

Baars, S; Bernardes, E.; Elwick, A.; Malortie, A.; McAleavy, T.; McInerney, L.; Menzies, L.; & Riggall, A. (2014) “Lessons from London’s schools: Investigating the success,” CfBT report.

Blanden, J.; Greaves, E.; Gregg, P.; Macmillan, L. & Sibieta, L. (2015) “Understanding the improved performance of disadvantaged pupils in London.” LSE/CASE working paper no. 21.

Burgess, S. (2014) “Understanding the success of London’s schools,” CMPO Working paper, no 14/333.

Burgess, S.; Greaves, E.; Vignoles, A. & D. Wilson (2014) “What Parents Want: School Preferences and School Choice,” *The Economic Journal*, pp.1—16.

Cantle, T. (2001) “Community Cohension: A report of the independent review team,” Home Office report.

Cantle, T. and Kaufmann, E. (2016) “Is segregation increasing in the UK?” Open Democracy Report.

Cook, C. (2013) “London School children perform the best”. *Financial Times*, 13/1/2013.

Epple, D. & Romano, R. (2012) “Economic modelling and analysis of educational vouchers.” *Annual Review of Economics*, 4 (1) pp. 159-183.

Gale, D. and Shapley, L. (1962) “College admissions and the stability of marriage,” *The American Mathematical Monthly,* 69 pp. 9-15.

Hastings, J.; Kane, T. & D. Staiger (2006) “Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery,” NBER Working Paper no. 12145.

iCoCo Foundation; SchoolDash; & The Challenge (2017) Understanding school Segregation in England. Technical report.

Ouseley, H. (2001) “Community pride not prejudice – making diversity work in Bradford – The Ouseley Report,” Report by Sir Herman Ouseley.

Weldon, M. and A. Titman (2015) “Estimation of Two-Sided Choice Models: An application to public school choice,” in *Proceedings of the 30th International Workshop on Statistical Modelling Vol. 1.* Eds. H Friedl and H Wagner, pp. 387—391.

Appendix 1: Estimated discrete choice models

Table 8: Estimates of model A for all locations

	London			Birmingham			Manchester			Pennines		
Coefficient	Est.	St. Err.		Est.	St. Err.		Est.	St. Err.		Est.	St. Err.	
AC5 (10 points)												
	0.239	(0.015)	***	0.241	(0.014)	***	0.350	(0.013)	***	0.262	(0.013)	***
x black	0.244	(0.022)	***	0.081	(0.023)	***	0.222	(0.038)	***	0.067	(0.050)	
x other	0.133	(0.022)	***	-0.039	(0.025)		0.044	(0.028)		0.025	(0.035)	
x South Asian	0.295	(0.022)	***	0.096	(0.015)	***	0.164	(0.022)	***	-0.091	(0.015)	***
x Pupil Premium	-0.235	(0.023)	***	-0.269	(0.016)	***	-0.307	(0.015)	***	-0.277	(0.016)	***
x PP & black	0.075	(0.032)	*	0.072	(0.029)	*	0.154	(0.048)	**	0.040	(0.068)	
x PP & other	0.102	(0.034)	**	0.085	(0.035)	*	0.194	(0.045)	***	-0.033	(0.060)	
x PP & South Asian	0.086	(0.037)	*	0.013	(0.022)		0.077	(0.033)	*	0.079	(0.025)	**
x KS2: High	0.505	(0.019)	***	0.341	(0.013)	***	0.259	(0.015)	***	0.295	(0.014)	***
x KS2: Low	-0.212	(0.013)	***	-0.213	(0.011)	***	-0.149	(0.014)	***	-0.111	(0.013)	***
Distance (km)												
	-0.537	(0.018)	***	-0.445	(0.017)	***	-0.547	(0.016)	***	-0.521	(0.016)	***
x black	0.136	(0.021)	***	0.125	(0.023)	***	0.060	(0.034)		-0.033	(0.049)	
x other	0.060	(0.023)	**	0.138	(0.027)	***	0.140	(0.032)	***	-0.028	(0.042)	
x South Asian	0.240	(0.020)	***	0.237	(0.017)	***	0.145	(0.024)	***	0.014	(0.020)	
x Pupil Premium	-0.051	(0.017)	**	-0.078	(0.017)	***	-0.136	(0.021)	***	-0.087	(0.020)	***
x KS2: High	0.248	(0.018)	***	0.073	(0.017)	***	0.099	(0.020)	***	0.133	(0.020)	***
x KS2: Low	-0.111	(0.020)	***	-0.124	(0.020)	***	-0.043	(0.022)		-0.031	(0.021)	
log(Distance)												
	-0.675	(0.043)	***	-1.139	(0.044)	***	-0.605	(0.042)	***	-0.605	(0.043)	***
x black	0.057	(0.053)		0.176	(0.061)	**	0.200	(0.089)	*	0.135	(0.121)	
x other	-0.017	(0.055)		0.005	(0.074)		-0.185	(0.087)	*	0.140	(0.109)	
x South Asian	-0.594	(0.051)	***	-0.256	(0.045)	***	-0.240	(0.065)	***	-0.211	(0.052)	***
x Pupil Premium	-0.051	(0.041)		0.029	(0.042)		0.115	(0.051)	*	0.009	(0.051)	
x KS2: High	-0.321	(0.055)	***	0.013	(0.051)		-0.188	(0.056)	***	-0.172	(0.058)	**
x KS2: Low	0.085	(0.044)		0.122	(0.049)	*	-0.063	(0.054)		-0.024	(0.053)	
Ofsted x local deprivation (IMD) fixed effects												
Ofsted: Outst'g & IMD: Hi	-0.011	(0.036)		0.067	(0.033)	*	0.529	(0.055)	***	0.129	(0.057)	*
Ofsted: Outs'ng & IMD: Mid	-0.017	(0.031)		0.302	(0.032)	***	0.283	(0.029)	***	0.229	(0.030)	***
Ofsted: Outs'ng & IMD: Lo	0.279	(0.043)	***	0.606	(0.039)	***	-0.008	(0.039)		-0.406	(0.047)	***
Ofsted: Good & IMD: Hi	-0.367	(0.034)	***	-0.269	(0.030)	***	-0.222	(0.032)	***	-0.501	(0.029)	***
Ofsted: Good & IMD: Mid[1]	0.000			0.000			0.000			0.000		
Ofsted: Good & IMD: Lo	0.349	(0.045)	***	0.455	(0.041)	***	0.337	(0.032)	***	0.378	(0.032)	***
Ofsted: < Good & IMD: Hi	-0.600	(0.050)	***	-0.641	(0.036)	***	0.084	(0.036)	*	-0.772	(0.032)	***
Ofsted: < Good & IMD: Mid	-0.350	(0.051)	***	-0.191	(0.032)	***	-0.027	(0.026)		-0.358	(0.027)	***
Ofsted: < Good & IMD: Lo	-0.281	(0.105)	**	0.025	(0.070)		0.315	(0.057)	***	0.088	(0.061)	
Faith school fixed effects												
Faith: Church of England	-0.077	(0.031)	*	-0.006	(0.039)		-0.145	(0.029)	***	-0.323	(0.030)	***
Faith: Roman Catholic	-0.089	(0.027)	***	0.056	(0.026)	*	-0.037	(0.022)		-0.077	(0.024)	**
# Pupils	21,890	[2]		28,373			25,891			23,363		
# Schools	435			175			157			150		
Pseudo R-squared	0.439			0.520			0.526			0.553		

[1] Coefficient is fixed at zero. [2] A random sample was taken from London's cohort of 78,000. Key: * p<0.05; ** p<0.01; *** p<0.001

Table 9: Estimates of model B for all locations

Coefficient	London Est.	St. Err.		Birmingham Est.	St. Err.		Manchester Est.	St. Err.		Pennines Est.	St. Err.	
AC5 predicted (10 points)												
	0.270	(0.017)	***	0.291	(0.016)	***	0.329	(0.015)	***	0.260	(0.014)	***
x black	0.252	(0.025)	***	0.058	(0.026)	*	0.216	(0.044)	***	0.109	(0.053)	*
x other	0.167	(0.027)	***	-0.037	(0.030)		0.058	(0.032)		0.016	(0.038)	
x South Asian	0.284	(0.026)	***	0.070	(0.018)	***	0.181	(0.025)	***	-0.119	(0.016)	***
x Pupil Premium	-0.331	(0.031)	***	-0.352	(0.021)	***	-0.319	(0.020)	***	-0.280	(0.018)	***
x PP & black	0.145	(0.040)	***	0.135	(0.034)	***	0.176	(0.057)	**	0.060	(0.074)	
x PP & other	0.175	(0.042)	***	0.121	(0.041)	**	0.172	(0.054)	**	-0.033	(0.068)	
x PP & South Asian	0.174	(0.046)	***	0.081	(0.027)	**	0.059	(0.040)		0.046	(0.029)	
x KS2: High	0.556	(0.020)	***	0.373	(0.013)	***	0.331	(0.017)	***	0.325	(0.015)	***
x KS2: Low	-0.279	(0.016)	***	-0.274	(0.013)	***	-0.181	(0.017)	***	-0.143	(0.014)	***
AC5 residual												
	0.159	(0.026)	***	0.194	(0.021)	***	0.401	(0.024)	***	0.227	(0.026)	***
x black	0.212	(0.038)	***	0.058	(0.040)		0.255	(0.079)	**	-0.147	(0.112)	
x other	0.085	(0.037)	*	-0.116	(0.043)	**	-0.031	(0.063)		0.087	(0.079)	
x South Asian	0.317	(0.039)	***	0.066	(0.027)	*	0.095	(0.046)	*	0.029	(0.031)	
x Pupil Premium	-0.093	(0.040)	*	-0.172	(0.025)	***	-0.301	(0.029)	***	-0.257	(0.033)	***
x PP & black	0.009	(0.054)		0.055	(0.051)		0.105	(0.099)		0.042	(0.146)	
x PP & other	0.055	(0.057)		0.185	(0.063)	**	0.327	(0.100)	**	-0.065	(0.118)	
x PP & South Asian	-0.013	(0.061)		0.037	(0.040)		0.182	(0.070)	**	0.109	(0.049)	*
x KS2: High	0.083	(0.036)	*	0.000	(0.023)		0.024	(0.030)		0.068	(0.033)	*
x KS2: Low	-0.088	(0.021)	***	-0.097	(0.019)	***	-0.091	(0.026)	***	-0.027	(0.026)	
Distance (km)												
	-0.539	(0.018)	***	-0.443	(0.017)	***	-0.546	(0.016)	***	-0.522	(0.016)	***
x black	0.133	(0.022)	***	0.122	(0.023)	***	0.058	(0.034)		-0.031	(0.049)	
x other	0.056	(0.023)	*	0.135	(0.027)	***	0.137	(0.032)	***	-0.031	(0.042)	
x South Asian	0.241	(0.021)	***	0.231	(0.017)	***	0.144	(0.024)	***	0.015	(0.020)	
x Pupil Premium	-0.049	(0.017)	**	-0.078	(0.017)	***	-0.136	(0.021)	***	-0.086	(0.020)	***
x KS2: High	0.231	(0.018)	***	0.077	(0.017)	***	0.094	(0.021)	***	0.134	(0.020)	***
x KS2: Low	-0.102	(0.020)	***	-0.124	(0.020)	***	-0.041	(0.022)		-0.033	(0.021)	
log(Distance)												
	-0.673	(0.043)	***	-1.142	(0.044)	***	-0.608	(0.042)	***	-0.605	(0.044)	***
x black	0.048	(0.053)		0.166	(0.061)	**	0.199	(0.089)	*	0.111	(0.121)	
x other	-0.021	(0.055)		0.005	(0.075)		-0.180	(0.087)	*	0.145	(0.109)	
x South Asian	-0.601	(0.052)	***	-0.258	(0.045)	***	-0.245	(0.065)	***	-0.219	(0.052)	***
x Pupil Premium	-0.043	(0.041)		0.044	(0.042)		0.119	(0.051)	*	0.009	(0.051)	
x KS2: High	-0.323	(0.055)	***	-0.031	(0.051)		-0.189	(0.056)	***	-0.165	(0.058)	**
x KS2: Low	0.082	(0.044)		0.136	(0.049)	**	-0.060	(0.054)		-0.019	(0.053)	
Ofsted x local deprivation fixed effects												
Ofsted: Outs'ng & IMD: Hi	0.059	(0.038)		0.044	(0.034)		0.530	(0.056)	***	0.108	(0.057)	
Ofsted: Outs'ng & IMD: Mid	0.003	(0.031)		0.269	(0.032)	***	0.289	(0.030)	***	0.256	(0.030)	***
Ofsted: Outs'ng & IMD: Lo	0.281	(0.044)	***	0.651	(0.039)	***	-0.010	(0.039)		-0.390	(0.050)	***
Ofsted: Good & IMD: Hi	-0.346	(0.035)	***	-0.272	(0.031)	***	-0.242	(0.033)	***	-0.544	(0.030)	***
Ofsted: Good & IMD: Mid[1]	0.000			0.000			0.000			0.000		
Ofsted: Good & IMD: Lo	0.325	(0.045)	***	0.461	(0.041)	***	0.348	(0.032)	***	0.391	(0.033)	***
Ofsted: < Good & IMD: Hi	-0.588	(0.050)	***	-0.654	(0.037)	***	0.083	(0.036)	*	-0.801	(0.032)	***
Ofsted: < Good & IMD: Mid	-0.373	(0.051)	***	-0.207	(0.032)	***	-0.026	(0.026)		-0.365	(0.028)	***
Ofsted: < Good & IMD: Lo	-0.361	(0.105)	***	0.047	(0.071)		0.315	(0.057)	***	0.110	(0.062)	
Faith school fixed effects												
Faith: Church of England	-0.091	(0.031)	**	-0.053	(0.040)		-0.132	(0.030)	***	-0.328	(0.030)	***
Faith: Roman Catholic	-0.080	(0.027)	**	0.022	(0.026)		-0.030	(0.022)		-0.072	(0.024)	**
# Pupils	21,890	[2]		28,373			25,891			23,363		
# Schools	435			175			157			150		
Pseudo R-squared	0.441			0.523			0.527			0.554		

[1] Coefficient is fixed at zero. [2] A random sample was taken from London's cohort of 78,000. Key: * p<0.05; ** p<0.01; *** p<0.001

Appendix 2: Willingness to Travel by location

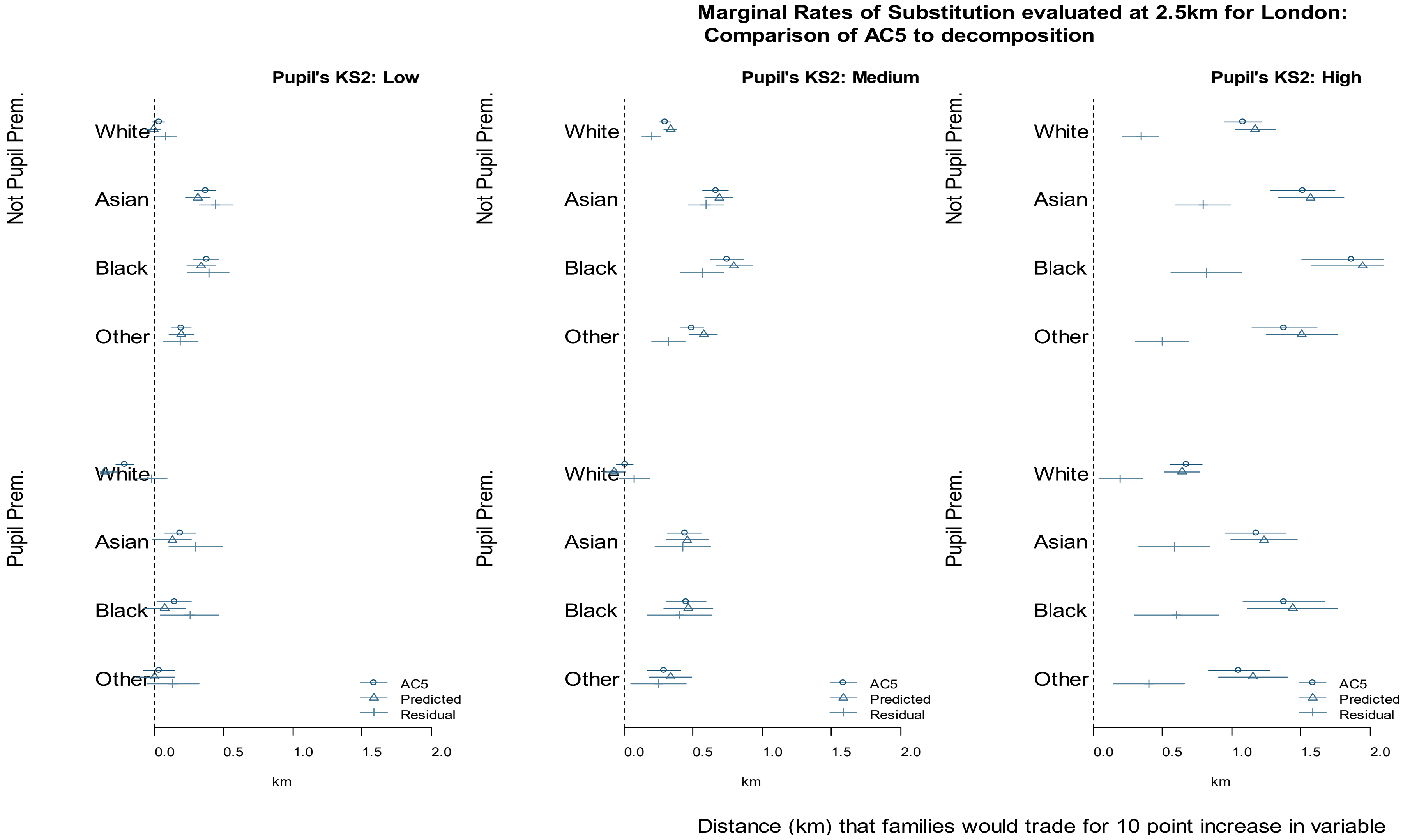

Figure 7: WTT for London: AC5; AC5_predicted; AC5_residual

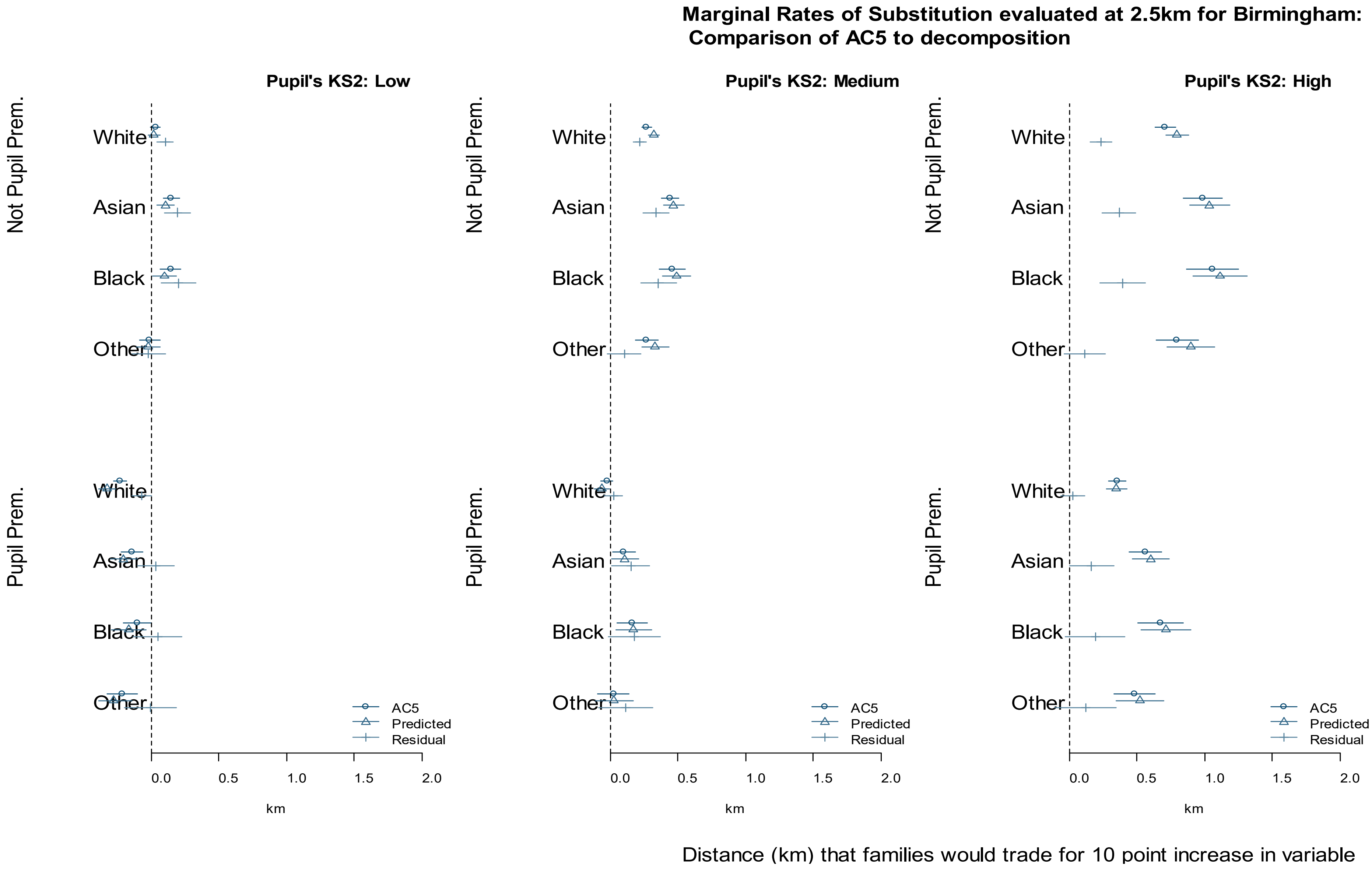

Figure 8: WTT for Birmingham: AC5; AC5_predicted; AC5_residual

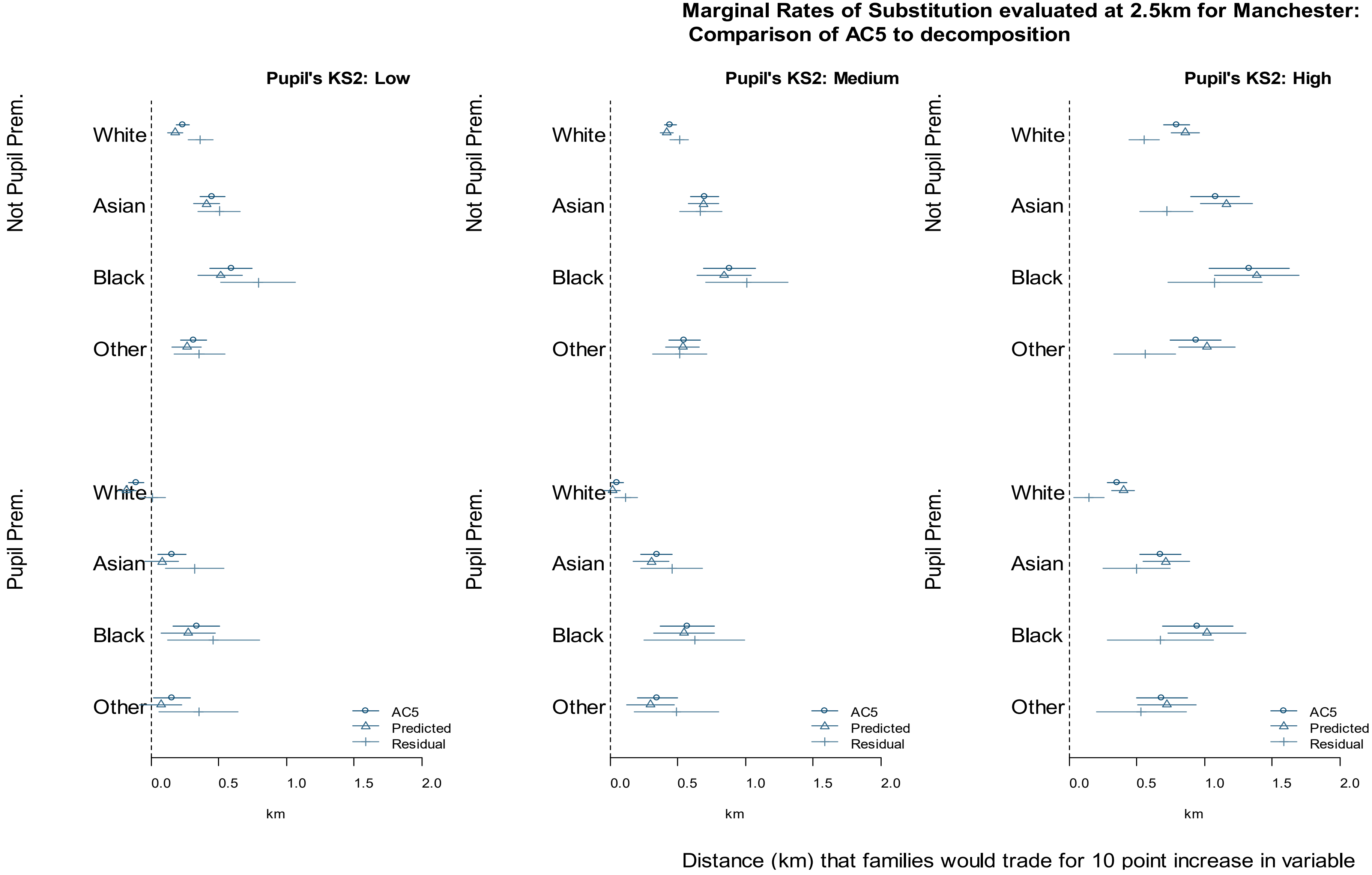

Figure 9: WTT for Manchester: AC5; AC5_predicted; AC5_residual

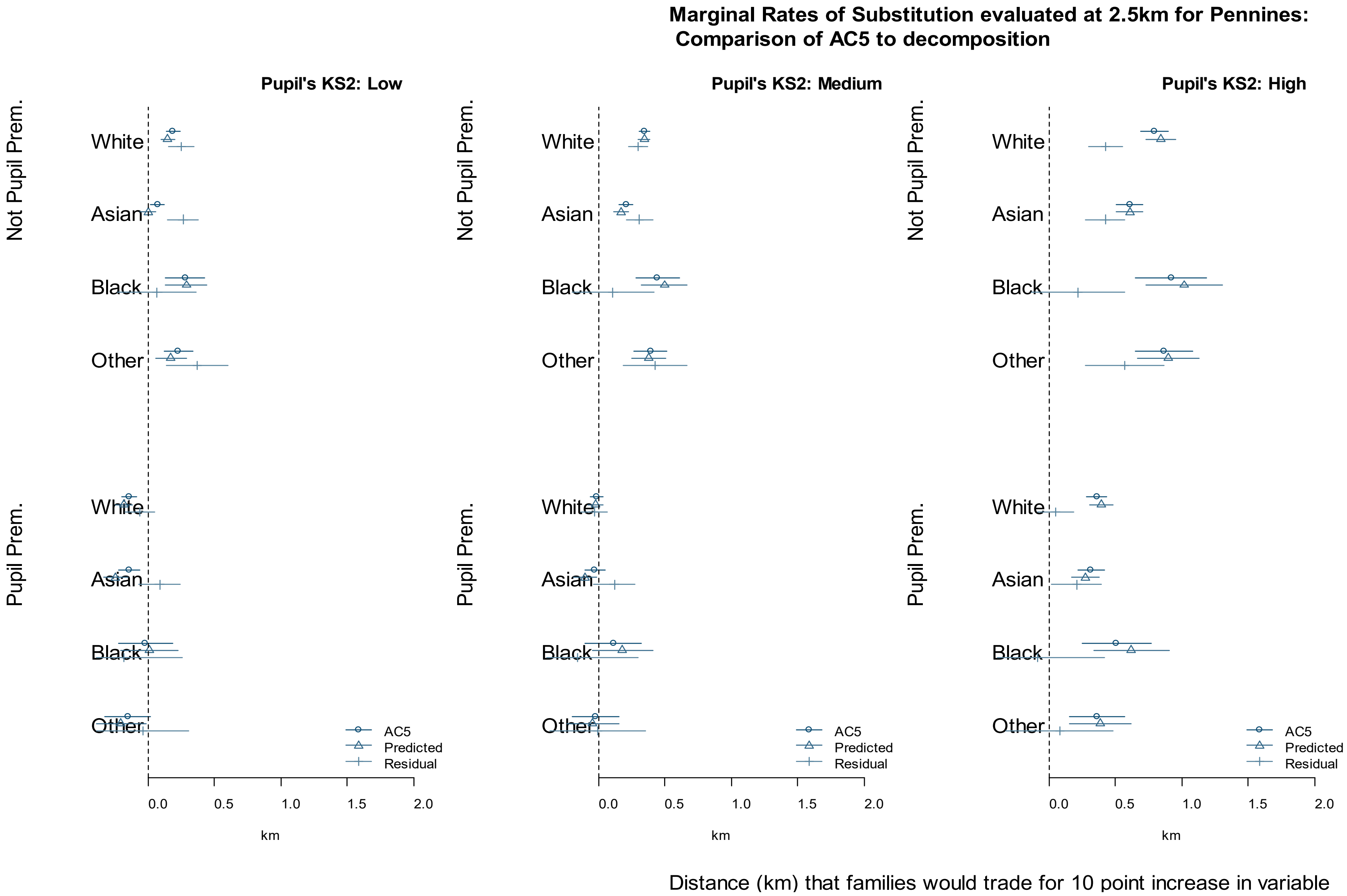

Figure 10: WTT for Pennines: AC5; AC5_predicted; AC5_residual

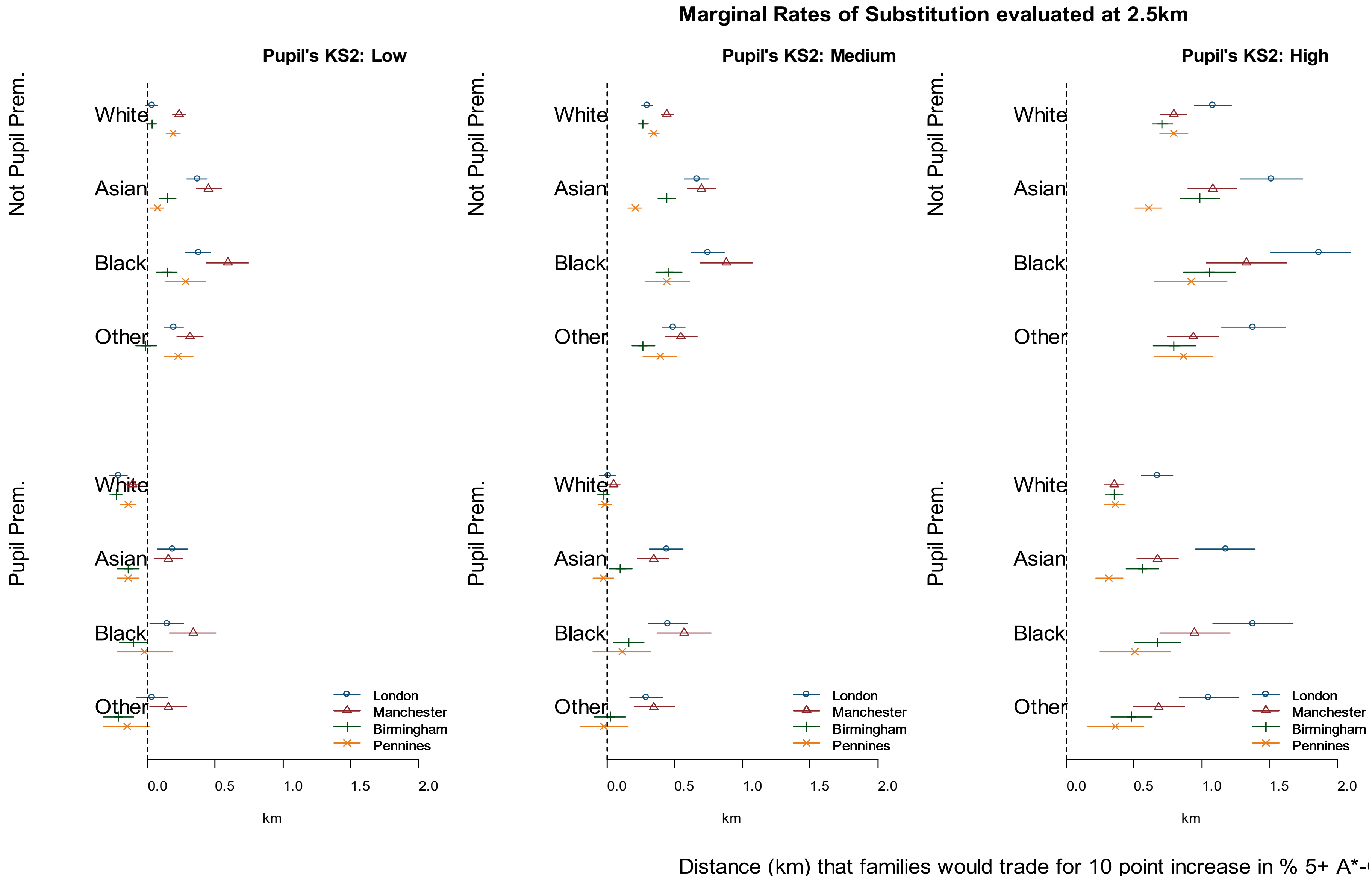

Figure 11: WTT on AC5 for all locations

Appendix 3: Admission results by radius and location

Table 10: Estimates of case-control logit for London

	Radius: 200m			Radius: 500m			Radius: 1000m		
Coefficient	Est.	St. Err.		Est.	St. Err.		Est.	St. Err.	
KS2: High	-0.291	(0.152)		-0.390	(0.096)	***	-0.233	(0.080)	**
KS2: Low	0.198	(0.102)		0.229	(0.069)	***	0.196	(0.060)	**
Pupil Premium	0.426	(0.099)	***	0.535	(0.066)	***	0.509	(0.057)	***
black	0.839	(0.144)	***	0.756	(0.094)	***	0.779	(0.078)	***
S. Asian	1.191	(0.176)	***	1.250	(0.113)	***	1.202	(0.095)	***
other ethnicity	0.860	(0.146)	***	0.728	(0.095)	***	0.657	(0.078)	***
SEN no st'mnt	0.132	(0.117)		0.052	(0.078)		0.098	(0.067)	
x Autonomous									
KS2: High	0.044	(0.183)		0.006	(0.132)		-0.119	(0.116)	
KS2: Low	-0.033	(0.127)		-0.096	(0.096)		-0.040	(0.089)	
Pupil Premium	-0.416	(0.124)	***	-0.519	(0.094)	***	-0.324	(0.085)	***
black	-0.534	(0.179)	**	-0.385	(0.131)	**	-0.401	(0.119)	***
S. Asian	-0.935	(0.213)	***	-1.001	(0.150)	***	-0.785	(0.135)	***
other ethnicity	-0.561	(0.180)	**	-0.441	(0.131)	***	-0.419	(0.116)	***
SEN no st'mnt	-0.082	(0.148)		0.068	(0.111)		0.030	(0.102)	
x Community									
KS2: High	0.340	(0.255)		0.648	(0.205)	**	0.250	(0.225)	
KS2: Low	-0.095	(0.183)		-0.153	(0.142)		-0.014	(0.153)	
Pupil Premium	-0.580	(0.175)	***	-0.564	(0.137)	***	-0.607	(0.146)	***
black	-0.549	(0.271)	*	-0.352	(0.203)		-0.627	(0.214)	**
S. Asian	-0.969	(0.290)	***	-0.839	(0.225)	***	-0.810	(0.246)	***
other ethnicity	-0.470	(0.261)		-0.338	(0.201)		-0.516	(0.233)	*
SEN no st'mnt	-0.058	(0.201)		0.036	(0.160)		0.054	(0.173)	
# Pupils	8,442			14,997			18,625		
ow. Cases	3,938			6,620			7,844		
ow. Controls	4,504			8,377			10,781		
# Schools	258			247			218		

Key: * p<0.05; ** p<0.01; *** p<0.001

Table 11: Estimates of case-control logit for Manchester & Birmingham

	Radius: 200m			Radius: 500m			Radius: 1000m		
Coefficient	Est.	St. Err.		Est.	St. Err.		Est.	St. Err.	
KS2: High	0.132	(0.161)		0.124	(0.117)		0.017	(0.099)	
KS2: Low	0.130	(0.156)		0.204	(0.110)		0.186	(0.095)	
Pupil Premium	0.780	(0.148)	***	0.750	(0.100)	***	0.789	(0.087)	***
black	0.448	(0.203)	*	0.595	(0.135)	***	0.808	(0.123)	***
S. Asian	0.899	(0.217)	***	1.193	(0.152)	***	1.427	(0.133)	***
other ethnicity	0.668	(0.242)	**	0.716	(0.165)	***	0.884	(0.138)	***
SEN no st'mnt	0.466	(0.177)	**	0.255	(0.124)	*	0.172	(0.111)	
x Autonomous									
KS2: High	-0.088	(0.221)		-0.050	(0.161)		-0.058	(0.147)	
KS2: Low	0.119	(0.200)		-0.070	(0.147)		-0.065	(0.135)	
Pupil Premium	-0.568	(0.188)	**	-0.595	(0.134)	***	-0.656	(0.123)	***
black	-0.581	(0.283)	*	-0.540	(0.202)	**	-0.579	(0.189)	**
S. Asian	-0.627	(0.276)	*	-0.949	(0.196)	***	-1.329	(0.183)	***
other ethnicity	-0.733	(0.340)	*	-0.528	(0.243)	*	-0.679	(0.223)	**
SEN no st'mnt	-0.248	(0.229)		-0.130	(0.171)		-0.112	(0.156)	
x Community									
KS2: High	-0.712	(0.304)	*	-0.381	(0.268)		-0.407	(0.278)	
KS2: Low	-0.568	(0.294)		-0.608	(0.255)	*	-0.548	(0.262)	*
Pupil Premium	-1.003	(0.266)	***	-0.758	(0.228)	***	-0.753	(0.254)	**
black	-0.193	(0.508)		0.094	(0.401)		-0.149	(0.453)	
S. Asian	-0.793	(0.477)		-1.345	(0.341)	***	-1.582	(0.339)	***
other ethnicity	-0.431	(0.580)		-0.639	(0.433)		-1.010	(0.469)	*
SEN no st'mnt	-0.424	(0.332)		-0.482	(0.290)		-0.873	(0.326)	**
# Pupils	3,735			7,152			9,078		
ow. Cases	1,621			2,784			3,156		
ow. Controls	2,114			4,368			5,922		
# Schools	158			165			152		

Key: * p<0.05; ** p<0.01; *** p<0.001

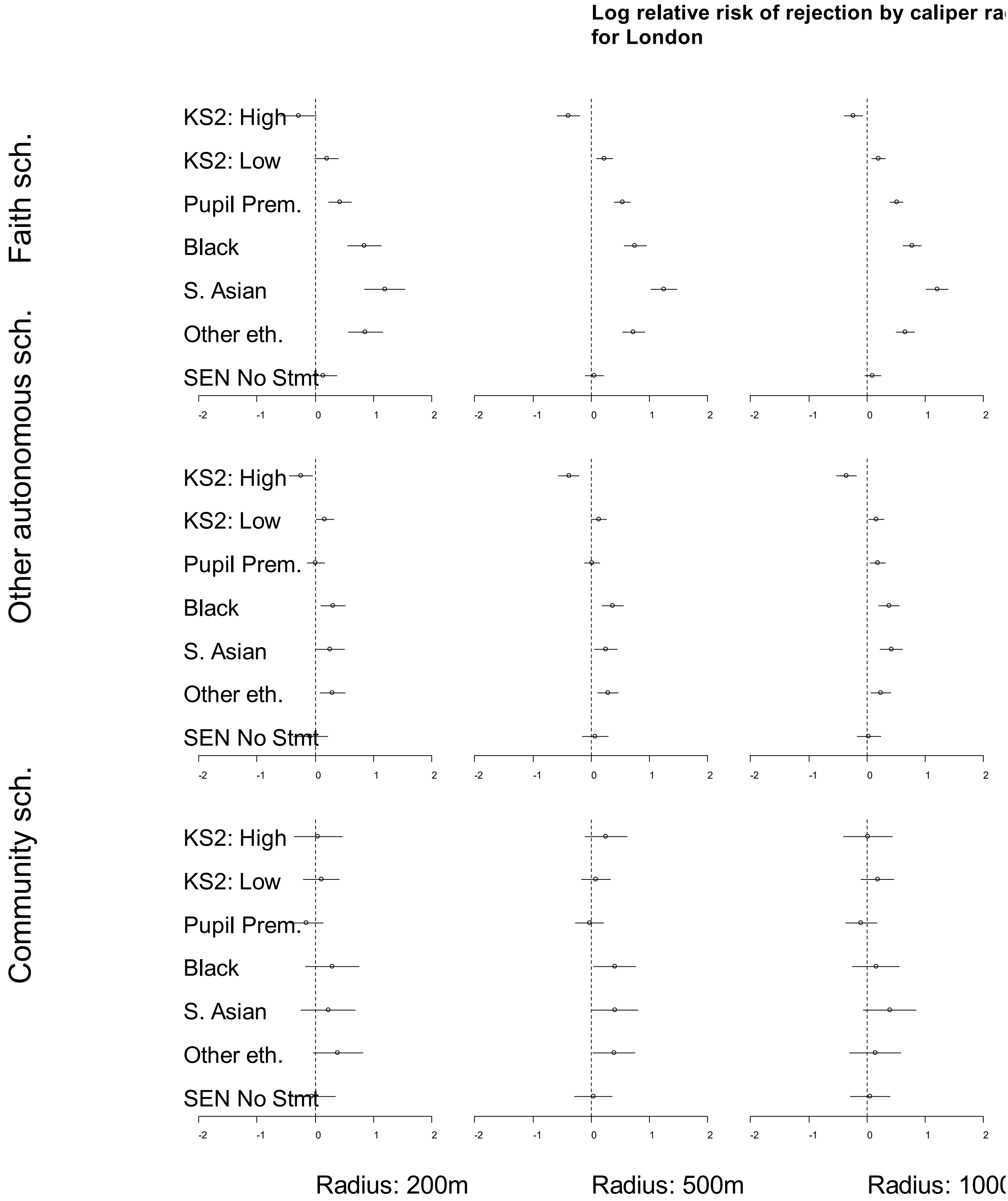

Figure 12: Estimated coefficients from the matched logit model of non-admission for London. Comparison of matching radii.

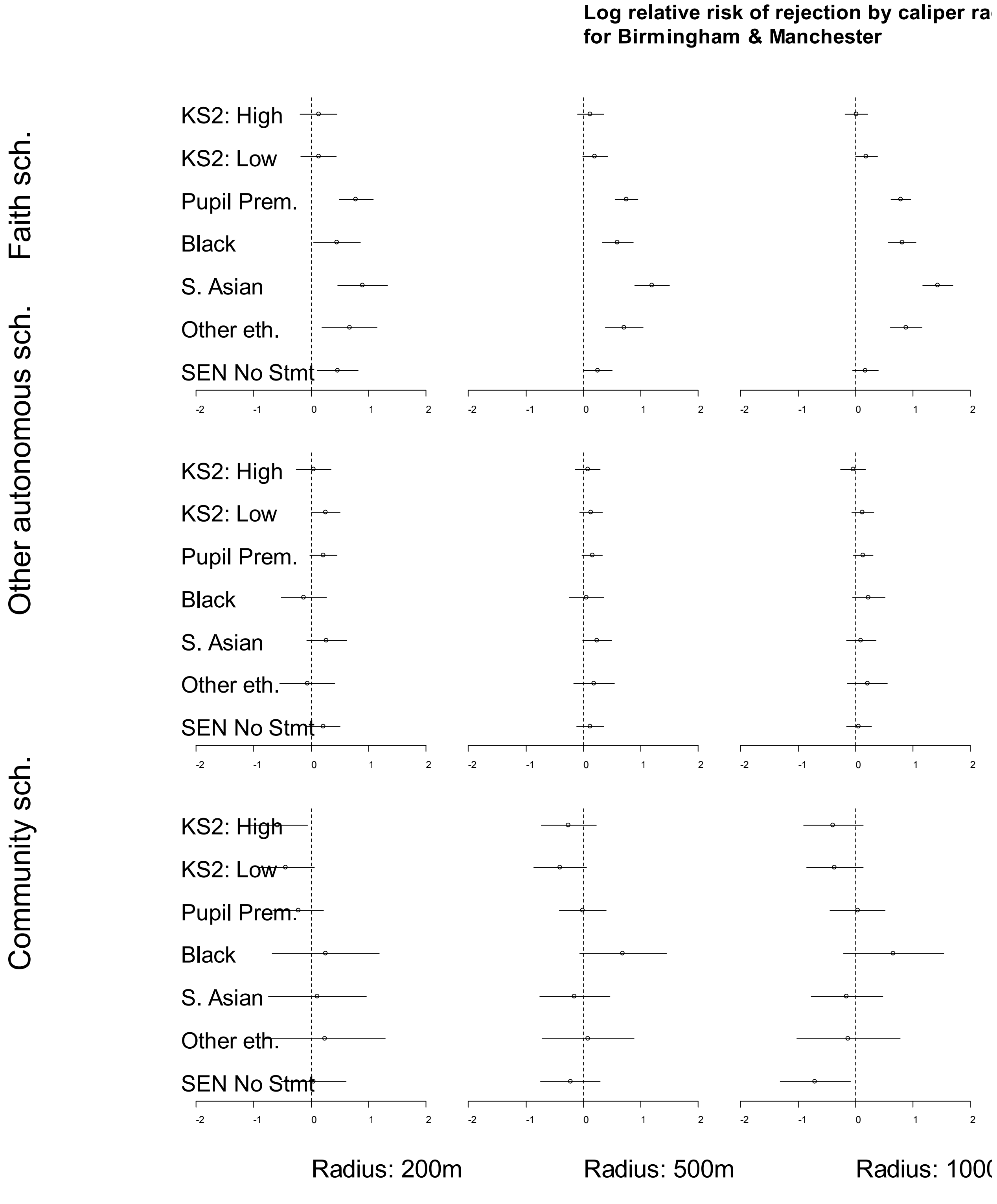

Figure 13: Estimated coefficients from the matched logit model of non-admission for Manchester & Birmingham. Comparison of matching radii.

Appendix 4: The multinomial logit discrete choice model

As described in section 5.1, the association between observed characteristics of families and schools, and the probability of a particular family choosing a school can be modelled by invoking a latent continuous variable that is usually known as *utility*.

The utility of a family i for a school j can be written as

$$U_{ij} = v\left(X_i, Z_j; \beta\right) + \varepsilon_{ij}$$

where the regression function $v(\cdot)$ represents the part of utility that is predictable from observed characteristics X_i and Z_j with coefficients β. ε_{ij} is the idiosyncratic random term. The model is fitted to data by estimating the coefficients β.

The form of the model depends on the random element, which is assumed to follow a particular distribution. If the random element is assumed to follow a Gumbel distribution, then the resulting model is the multinomial logit model, and the regression model $v(\cdot)$ is transformed into a model of the probability of choice by

$$P(i \text{ chooses } j) = \frac{e^{v\left(X_i, Z_j; \beta\right)}}{\sum_{h=1}^{J} e^{v\left(X_i, Z_h; \beta\right)}} .$$

The coefficients, β, of this model can be estimated using maximum likelihood methods.

In the study two models for the utility prediction function $v(\cdot)$ are specified and estimated. The first model is specified as

$$\begin{aligned} U_{ij} = \beta_1 d_{ij} + \beta_2 \log d_{ij} + \beta_3 AC5_j + PP_i\left(\beta_4 d_{ij} + \beta_5 \log d_{ij} + \beta_6 AC5_j\right) \\ + black_i\left(\beta_7 d_{ij} + \beta_8 \log d_{ij} + \beta_9 AC5_j\right) \\ + Asian_i\left(\beta_{10} d_{ij} + \beta_{11} \log d_{ij} + \beta_{12} AC5_j\right) \\ + other_i\left(\beta_{13} d_{ij} + \beta_{14} \log d_{ij} + \beta_{15} AC5_j\right) \\ + KS2lo_i\left(\beta_{16} d_{ij} + \beta_{17} \log d_{ij} + \beta_{18} AC5_j\right) \\ + KS2hi_i\left(\beta_{19} d_{ij} + \beta_{20} \log d_{ij} + \beta_{21} AC5_j\right) \\ + PP_i AC5_j(\beta_{22} black_i + \beta_{23} Asian_i + \beta_{24} other_i) + \sum_{k=1}^{3}\sum_{l=1}^{3} \beta_{kl} Ofsted_{jk} IMD_{jl} \\ + \beta_{25} CofE_j + \beta_{26} RC_j + \varepsilon_{ij} . \end{aligned}$$

$$\varepsilon_{ij} \sim \text{Gumbel} .$$

In this notation, each of the ethnic and KS2 groups, and Pupil Premium, have an indicator variable which takes a value of zero if pupil i does not belong to that group, and one if she does. These indicator variables mean that certain parts of the model are only relevant to certain groups. For example, coefficients β_4 to β_6 only apply to pupils eligible for Pupil Premium, otherwise that part of the model is zero. The variable d_{ij} denotes distance, in km, and log denotes the natural logarithm.

The second model is specified as

$$
\begin{aligned}
U_{ij} = \beta_1 d_{ij} + \beta_2 \log d_{ij} + \beta_3 AC5\_pred_j + \beta_4 AC5\_resid_j \\
+ PP_i\left(\beta_5 d_{ij} + \beta_6 \log d_{ij} + \beta_7 AC5\_pred_j + \beta_8 AC5\_resid_j\right) \\
+ black_i\left(\beta_9 d_{ij} + \beta_{10} \log d_{ij} + \beta_{12} AC5\_pred_j + \beta_{13} AC5\_resid_j\right) \\
+ Asian_i\left(\beta_{14} d_{ij} + \beta_{15} \log d_{ij} + \beta_{16} AC5\_pred_j + \beta_{17} AC5\_resid_j\right) \\
+ other_i\left(\beta_{18} d_{ij} + \beta_{19} \log d_{ij} + \beta_{20} AC5\_pred_j + \beta_{21} AC5\_resid_j\right) \\
+ KS2lo_i\left(\beta_{22} d_{ij} + \beta_{23} \log d_{ij} + \beta_{24} AC5\_pred_j + \beta_{25} AC5\_resid_j\right) \\
+ KS2hi_i\left(\beta_{26} d_{ij} + \beta_{27} \log d_{ij} + \beta_{28} AC5\_pred_j + \beta_{29} AC5\_resid_j\right) \\
+ PP_i AC5\_pred_j(\beta_{30} black_i + \beta_{31} Asian_i + \beta_{32} other_i) \\
+ PP_i AC5\_resid_j(\beta_{33} black_i + \beta_{34} Asian_i + \beta_{35} other_i) \\
+ \sum_{k=1}^{3}\sum_{l=1}^{3} \beta_{kl} Ofsted_{jk} IMD_{jl} + \beta_{36} CofE_j + \beta_{37} RC_j + \varepsilon_{ij} \,.
\end{aligned}
$$

$$
\varepsilon_{ij} \sim \text{Gumbel} \,.
$$

Willingness to Travel

The Willingness to Travel for academic performance, otherwise known as the *marginal rate of substitution*, can be interpreted as the infinitesimal increase in distance to a school that would exactly compensate for an infinitesimal improvement in test scores, leaving a family's probability of choosing a school unchanged. It also serves as a linear approximation to the increase in distance that a parent will be "willing" (in the economic sense described above) to exchange for a unit improvement in performance. Since, in the model, AC5 is measured in units equivalent to 10 percentage points, $WTT_{AC5,d}$ therefore approximately measures the distance, in km, that a parent will travel for a 10 point improvement in AC5.

The Willingness to Travel is estimated by

$$
WTT_{AC5,d} = \frac{\partial U/\partial AC5}{\partial U/\partial d} = \frac{\hat{\beta}^{(AC5)}}{\hat{\beta}^{(d)} + \frac{1}{d}\hat{\beta}^{(\log d)}} \,.
$$

In the above expression, the estimated composite coefficients aggregate the model coefficients that are relevant for the demographic characteristics of the pupil. For example, in the first model above, $\hat{\beta}^{(AC5)}$ for a South Asian, Pupil Premium pupil would be constructed as $\hat{\beta}_3 + \hat{\beta}_6 + \hat{\beta}_{12} + \hat{\beta}_{23}$. Standard errors are constructed using the delta method, ignoring covariance terms.

Appendix 5: The conditional logit case-control model

In the case-control matched analysis of rejection, we wish to model the probability of non-admission to different types of school as a function of pupil characteristics: ability; disadvantage; ethnic group; and SEN status. Matching on location removes the important influence of residential location on school access probabilities, but it also distorts the proportions of admitted and unadmitted pupils in the sample. For example, using a one-to-one case-control design would fix the overall probability of non-admission *in the sample* to 0.5. The conditional logit model accounts for the matched sampling by treating each matched pair (or group, in the case of many-to-one matches) as a self-contained sub-sample or *stratum*, and modelling the probability of non-admission for each child within each stratum, conditional upon the fixed admission rate in the stratum.

The observed characteristics of child i are related to the probability of not being admitted by a regression function or linear predictor,

$$\eta_i = \sum_{k=1}^{K} \gamma_k x_{ik} \,,$$

where γ are the coefficients to be estimated , and k indexes the different observed pupil characteristics, $x.$ The conditional logit probability depends on the number of cases and controls in each stratum. In a one-to-many design where, within each stratum *t,* one case is matched to $H_t - 1$ controls, the probability is

$$P(i \text{ is not admitted in stratum } t) = \frac{e^{\eta_i}}{\sum_{h=1}^{H_t} e^{\eta_h}}.$$

In this case, the conditional logit model turns out to be mathematically identical to the multinomial logit discrete choice model. This makes sense: the discrete choice model represents the probability of choice of one school from a fixed number of options; whereas the case-control model represents the probability of "rejection" of one child, given a fixed number of children in the stratum. The many-to-one design, where $H_t - 1$ cases are matched to one control, is similar, but the denominator is a more complex expression that is calculated iteratively.

The method presented in this report allows both many-to-one and one-to-many matchings. Many-to-many matched designs, although possible, are more computationally demanding to estimate so are not used in the study.

The regression function used in the analysis includes indicator variables for child characteristics, and interacts these with indicators of school type, to estimate separate rejection probabilities for the three school types:

$$\eta_i = \gamma_1 KS2hi_i + \gamma_2 KS2lo_i + \gamma_3 PP_i + \gamma_4 black_i + \gamma_5 Asian_i + \gamma_6 Other_i + \gamma_7 SEN_i$$
$$+ \ Autonomous_i(\gamma_8 KS2hi_i + \gamma_9 KS2lo_i + \gamma_{10} PP_i + \gamma_{11} black_i + \gamma_{12} Asian_i$$
$$+ \gamma_{13} Other_i + \gamma_{14} SEN_i)$$
$$+ Community_i(\gamma_{15} KS2hi_i + \gamma_{16} KS2lo_i + \gamma_{17} PP_i + \gamma_{18} black_i + \gamma_{19} Asian_i$$
$$+ \gamma_{20} Other_i + \gamma_{21} SEN_i) \ .$$

Interpretation of coefficients

To ease interpretation of the estimated coefficients, $\hat{\gamma}$ they are transformed into a particular type of non-admission probability. To motivate this transformation, we consider a hypothetical adversarial situation where two children are assessed for the last remaining seat at a school. One of the children possesses the characteristic relating to the coefficient (say, Pupil Premium). We call this the child of interest. The other child has none of the characteristics and is therefore white, medium KS2, non-Pupil Premium and without Special Educational Needs. We call this the baseline child. The probability that the child of interest *is not admitted*, while the baseline child is admitted in this hypothetical situation is given by

$$P(i \text{ is rejected vs. baseline child}) = \frac{e^{\hat{\gamma}}}{1 + e^{\hat{\gamma}}} \, .$$

Reference: DFE-RR799

ISBN: 978-1-78105-911-1

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: allan.little@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications