cemmap

centre for microdata methods and practice

Teacher-pupil sorting, learning, and inequality

Ivan Olszak

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP11/25

Economic and Social Research Council

Teacher-pupil sorting, learning, and inequality

Ivan Olszak (UCL/IFS)

May 2024

Abstract

This paper uses a structural, two-sided model of the education system to study the interactions between parents' school choices and teachers' labour supply decisions in the context of secondary education in England. I find that more affluent households put more weight on school performance when applying for school places, and teachers tend to prefer working for schools with children from more affluent families. These preferences generate sorting effects where children from more disadvantaged households tend to be taught by less experienced and less effective teachers, which increases inequality in learning outcomes. The simulation of policy counterfactuals sows that funding for schools serving disadvantaged communities would have to increase very substantially to counter these sorting effects and reduce inequality.

1 Introduction

Over the past three decades, many governments have introduced systems of school choice, whereby parents can apply for places at their preferred schools instead of seeing their children assigned to the local school. The main objective of these policies is to improve pupil outcomes by allowing households located close to poorly-performing schools to access better schools. However, to evaluate the equilibrium effects of school choice, it is critical to understand how schools - the 'supply side' in the education market - respond to parents' decisions. Proponents of the policy often argue that schools are likely to respond to the introduction of school choice by exerting more effort and improving performance to attract pupils. This argument is plausible in education systems dominated by private, for-profit schools, but it is not clear that it applies in systems dominated by state-funded schools, where there is no profit incentive for management to increase school attendance.

In this paper I propose a different model of the school system which incorporates teachers' decisions of where to work. The premise is that teachers are likely to have preferences over the non-wage characteristics of teaching jobs, including the composition of a school's pupil body. These preferences are likely to affect the ability of different schools to recruit teachers, which, in turn, is likely to affect their performance and their attractiveness to households. Hence, in this model, pupil outcomes are shaped by the interactions between parents' and teachers' choices, which can create sorting effects over schools, and increase or decrease inequality in access to learning opportunities.

Suppose for example that relatively affluent families place more emphasis on school performance in their applications, and teachers prefer to work for schools with higher shares of pupils from affluent families. Then the introduction of school choice means that poorly-performing schools located in disadvantaged areas will lose pupils from more affluent families, which might increase their difficulties to recruit teachers, which might affect their performance, which in turn may affect their attractiveness to affluent families, etc. If this is correct, then it is no longer true that the introduction of school choice benefits everyone; instead it might reduce learning opportunities for the most disadvantaged pupils.

These interactions are also likely to play an important part in the design of school funding policies. Many education systems grant additional funding to schools located in disadvantaged areas with a view to reduce inequality in pupil attainment. The extra funding needed to reduce inequality by a given amount will depend not just on teachers' preferences, but also on equilibrium responses by pupils to changes in the teaching workforce.

I study these interactions in the context of secondary education in England. England has had a system of school choice since the 1980s, introduced under the belief that it would improve pupil attainment and school effectiveness.[1] Also, in England, 95% of pupils are educated in state-funded schools, one of the highest shares in the OECD, and this allows me to focus on sorting effects within the public sector, rather than between the public and private sectors.

To understand these interactions and model relevant policy counterfactuals, I build a structural, 'two-sided' model of the secondary school system in England. Parents have preferences over school attributes, which include distance to their home, performance metrics, and the proportion of disadvantaged pupils. Teachers also have preferences over school attributes, including the wages they offer and the proportion of disadvantaged pupils on their roll. Teachers are heterogeneous both in terms of their preferences and their effectiveness at improving pupils' test scores. The test scores of pupils depend on their socio-economic status, the composition of the pupil body of their school, and the composition of the teacher body of their school. Schools receive a budget allocation and set wages to hire the teacher body that maximise the test scores of the pupils on their rolls.

Unlike in many countries studied in recent literature (for example Peru in Ederer 2022, or France in Combe, Tercieux, and Terrier 2022), England does not have a centralised mechanism for allocating teachers to state-funded schools. That is, there is no central organization collecting teachers' applications and schools' preferences, and using a formal matching algorithm to allocate teachers to school. Instead, English schools simply advertise vacancies on their websites or on newspaper websites (eg the Times Educational Supplement, or Guardian Jobs), and teachers apply to vacancies and attend job interviews at schools. This presents a significant challenge for the analysis: the assignment mechanism is not observed, and there is no explicit information on teachers' and schools' preference rankings over the other side of the market. All that is observed is the result of this decentralized matching process, ie where teachers work,

[1] For example, in 2005 Prime Minister Tony Blair stated that 'parent choice can be a powerful driver of improved standards. Performance tables and inspections have given many parents the information that has enabled them to make objective judgements about a school's performance and effectiveness. This has been an important pressure on weaker schools to improve' (Department for Education White Paper 'Higher Standards, Better Schools For All:More choice for parents and pupils')

the terms of their employment, and the performance of schools. This decentralized matching process is also likely to be subject to various sources of frictions, for example if teachers are not perfectly aware of all vacancies, or if the vacancies advertised at any given time do not match their skill sets (eg a vacancy for an Maths teacher is not relevant for an English teacher).

I model these complex interactions using a model of job search, where teachers receive offers from schools at random rates, and move to a different job when the value of an offer exceeds the value of the job they hold. This approach has been used previously to model some aspects of the labour market and estimate workers preferences at a more macro level (see for example Sorkin 2018 and Moser and Morchio 2023). In this paper I enrich this approach to develop a detailed model of a specific labour market for a single occupation - teaching in secondary schools in England.

The estimation of the parameters in this structural model relies on rich administrative data coming from two different databases held by the UK Department for Education (DfE): data on pupils, their applications and their attainment at the end of secondary school comes from the National Pupils Database (NPD); and data on teachers, their characteristics and the conditions of their employment comes from the School Workforce Census (SWFC). These databases categorise pupils and their households according to whether they are eligible for Free School Meals (FSMs).

On the demand side, the analysis shows that more disadvantaged households put more weight on distance (ie face higher travel costs), and less weight on school performance and school composition. These patterns of heterogeneity imply that, in and of itself, the expansion of school choice may not necessarily reduce educational inequality, insofar as pupils from more affluent families are more likely to use the scheme to seek admission to better performing schools. This analysis also indicates that pupils' unobserved preferences for quality (which can be recovered from the mixed logit model) correlates with their attainment scores; in other words, pupils whose parents put more weight on school performance are also likely to do better in secondary school exams after controlling for their observable characteristics. This implies that measures of school value-added that control solely for observable pupil characteristics are imperfect proxies of true value-added.

On the supply side, the analysis shows that teachers tend to dislike working for schools with large shares of disadvantaged pupils, but there is significant heterogeneity among teachers in both preferences and effectiveness. One category of teachers appears to have a slightly stronger distaste for working in disadvantaged schools, but is also relatively more effective than other teachers in such schools. These teachers tend to be paid more than other teachers, especially by more disadvantaged schools, and are more likely to work in such schools.

Against this backdrop, a policy maker interested in reducing educational inequality might consider increasing the funding available to schools serving disadvantaged bodies of pupils. My simulations indicate that the effect of such policies would be limited: doubling the 'pupil premium' (the additional budget allowance granted to school for each disadvantaged pupil) from £1k to £2k per year would only reduce the mean attainment gap between affluent and disadvantaged pupils by 0.005 standard deviations. This is because, even at this increased level, the pupil premium only represents a small fraction of the funding available to schools, and also because differences in school value added only account for a share of the variance in educational attainment.

This work seeks to connect three strands of the literature on education and school choice. The first strand recovers estimates of household preferences using applications submitted in school choice mechanisms. In the UK context, this research has mostly used conditional logit models to estimate preferences under the assumption that applications truthfully reveal households' preferences (Burgess et al. 2015 and Weldon 2018). I build on this work by allowing for unobserved heterogeneity in preferences and instrumenting for endogenous school characteristics, in the spirit of the Industrial Organization literature on differentiated products markets.

The second strand seeks to recover estimates of teachers' preferences over school attributes and the corresponding 'compensating differentials' (the wage premium required to induce teachers to work in undesirable schools). This literature uses a variety of approaches, including hedonic regressions (Antos and Rosen 1975), search models (Bonhomme, Jolivet, and Leuven 2016), and matching models (Boyd et al. 2013, Ederer 2022, Combe, Tercieux, and Terrier 2022). Beyond the education-focused literature, recent research has used models of job search to investigate the role of job amenities and search frictions in explaining the career paths of different workers and their matching to firms (Sorkin 2018, Moser and Morchio 2023).

The third strand of literature seeks to evaluate the value added of schools and the effectiveness of teachers in raising educational attainment. I use information contained in households' applications to address the selection bias in the estimation of school value added, in a way similar to Dale and Krueger 2002 and, more recently, Abdulkadiroglu et al. 2020. In contrast to these papers, I use individual-level estimates of unobserved tastes for school performance to control for unobserved drivers of educational attainment. I also seek to understand how school value added is affected by equilibrium allocations in the teacher labour market, as in Stromme, Fu, and Biasi 2021 and Bates et al. 2022.

There is a small literature connecting some of these three themes to understand how school choice affects school performance and/or teachers allocation, primarily in contexts where private schools play a greater role. Tincani 2021 uses a structural model of the school system in Chile to study the sorting of pupils and teachers between the private and public sectors. Allende 2019 uses a structural model of the school system in Peru to examines how private schools choose the quality they offer depending on local competition and local market characteristics. In contrast, this paper is concerned with sorting effects within the public, not-for-profit sector. Like Allende, Gallego, and Neilson 2019, I am interested in understanding the equilibrium effect of school choice, but my model of the 'supply side' focuses on the labour supply decision of teachers as a key mechanism in shaping these interactions.

More generally, there is large body of literature that studies the spatial determinants of access to educational opportunities (eg Agostinelli, Luflade, and Martellini 2024), and the formulation of social interaction models with neighborhood and peer effects (eg Durlauf and Ioannides 2010). In my model, the socio-economic environment in which schools operate shape learning outcomes in two ways: a direct way, through interactions in the classroom, and an indirect way, through the labour supply decisions of teachers and their impact on school value added. The model presented in this paper can also be seen as an attempt to integrate an model of production into a broader analysis of a market that does not use prices to allocate resources, in the way suggested by Pakes 2021.

Mandatory disclaimer

This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. The analysis was carried out in the Secure Research Service, part of the Office for National Statistics

2 Context and institutions

In England, children transition from primary to secondary education at age 11. In the last year of primary school, parents are required to submit a list of secondary schools ranked in order of preference. They are allowed to list between 3 and 6 schools, depending on the Local Authority (LA) where they live, and are allowed to list schools outside their home LA. At the point of applying, parents are provided with a booklet containing information about local schools, their admission criteria, and some advice on how to fill their application. Parents are also encouraged to seek additional information from government websites and from the schools themselves.

Schools are required to publish the prioritization criteria they will use to rank applications if they are oversubscribed. With the exception of a small number of selective schools (known as 'grammar schools'), state schools are not allowed to prioritize pupils based on their ability. The structure of these admission policies typically involves a number of coarse priority groups which define a high-level ranking, and a continuous indicator used to break ties between applicants in the same priority group. Coarse priority groups can include children in care, children who have a sibling enrolled at the school, children who can demonstrate religious observance in a particular faith (for religious schools), and children of a particular sex (for single-sex schools). The continuous indicator used to break ties is typically the distance between a child's home and the school.

LAs collect parents' applications and schools' priority criteria, and allocate school places using a deferred-acceptance algorithm. If a child cannot be offered a place at any of the schools she applied to, the LA will typically allocate her to the nearest school with capacity.

After 5 years in secondary school, most pupils take exams to obtain their General Certificate of Primary Education (GCSE). The UK Department for Education (DfE) uses GCSE results and other data to compute various measures of attainment and value-added for secondary schools, which are then incorporated into 'performance tables'.

Schools are subject to a complex accountability framework which incorporates the quantitative metrics published in performance tables, but also wider contextual factors including the composition of their body of pupil. Until 2016, the headline measure of performance published in performance tables was the proportion of pupils who obtained grades of A to C in at least five GCSE subjects including English and Maths (a measure sometimes referred to as AC5EM). In 2016, the DfE introduced additional performance metrics including Progress 8, which is a simple measure of value added.[2] The headline measure of pupil composition is the percentage of pupils eligible for Free School Meals (FSMs). To be eligible for FSMs a child or their parent/carer must be in receipt of a qualifying benefit. In the remainder of this paper I sometimes refer to households with a child eligible for FSMs as 'disadvantaged' and to other households as 'affluent'.

The rules applicable to teacher pay and hiring vary across different types of schools in England. Schools that are structured as 'Academies' or 'Free Schools' have complete autonomy over their pay policies. 'Maintained schools', which include several categories of schools under various degrees of supervision by local authorities, are required to follow guidance set out by the DfE in the School Teachers Pay and Condition Document (STPCD). The STPCD sets out acceptable pay ranges for teachers that vary depending on roles (classroom teacher, leading practitioner, leadership group, etc) and the location of the school (inner London, outer London, Fringe, rest of England and Wales). Historically, teachers were recruited at the lowest point on the applicable pay range and their pay increased automatically with years of service. In 2013, the government introduced reforms giving maintained schools more flexibility to decide on starting salaries and requiring them to link pay progression to performance instead of years of service, such that maintained schools now also have substantial autonomy in setting their pay policies within the ranges set out in the STPCD. Maintained schools are also allowed to take account of local labour market conditions when setting wages. A 2017 report commissioned by the DfE found that schools had largely implemented these reforms, and had started to differentiate pay between teachers based on their assessed performance and local market conditions (DepartmentForEducation 2017). In summary, all schools have at least a degree of autonomy in setting wages to teachers, though most still have regard to central guidance.

During the period considered in this paper, the funding of each school was determined in two stages. In the first stage, the government allocated a Dedicated Schools Grant (DSG) to each LA, based primarily on the number of pupils on roll in the LA and its per-pupil allocation in the previous year. The baseline per-pupil allocation for each LA was set in 2005, taking into account factors such as deprivation, population sparsity, and area costs. In the second stage, each LA allocated funding from the DSG to individual schools based on a 'local funding formula'. These formulas had to reflect local deprivation indices, and could reflect additional factors such as the prior attainment of pupils or the number of children in care although this was not mandatory and there were substantial variations in practices between LAs. This system was generally considered unfair, in that DSGs reflected LAs' historical rather than current circumstances, and in 2018 the government introduced a National Funding Formula (NFF) with a view to re-aligning funding with individual school circumstances.

Schools also receive an additional allowance for each disadvantaged pupil on their roll, a system known as the 'Pupil Premium', which is currently £1,035 per pupil. A pupil qualifies for the Pupil Premium if she has been eligible for Free School Meals (FSMs) at any point in

[2] To calculate Progress 8 scores, the Department for Education puts all pupils nationally into prior attainment groups based on their test scores at the end of primary school. The Progress 8 score of an individual pupil is the difference between her test score at the end of secondary school and the mean test score at the end of primary school for her attainment category. The Progress 8 score for a school is then simply the average of the Progress 8 scores of all pupils enrolled at that school.

the last 6 years.[3]

3 Data and descriptive analysis

This paper combines several administrative data sources from the DfE. Data on pupils is obtained from the National Pupils Database (NPD). For each pupil who applied for a place in a state-funded secondary school in England for the 2014/15 academic year, I observe: the rank-ordered list of schools submitted at the point of application (in Autumn 2013), some individual characteristics (including their residential postcode and whether they were eligible for FSMs in February 2014), their test scores at the end of primary school (in Spring 2014), their test scores at the end of secondary school (in Spring 2019) and the school they were attending at that point.

Data on teachers is obtained from the School Workforce Census (SWFC). For each teacher who worked for a state-funded secondary school in England at some point over the period 2010-2019, I observe, for each year: the school where they worked, some individual characteristics (including qualification status and years of experience), and the main terms of their contracts (including wage, hours worked, role, etc).

Data on schools is obtained from the NPD and public sources (notably the DfE's Get Information about Schools website). For each state-funded secondary school in England, I observe: the location, key performance characteristics, and the proportion of pupils on their roll who is eligible for FSMs.

For computational reasons, I estimate and simulate my model on a subset of this data. For reasons that I detail in section 5.1, I focus on the West Midlands region. The West Midlands are a large, mostly urban area with 351 schools distributed over 14 LAs.

3.1 School applications

In 2014, 55,445 households applied for a place in secondary schools in the region (20.5% of which were eligible for FSMs), which corresponds to 11% of the total number of applicants in England in that year. Table 1 below shows the distribution of the number of schools ranked by pupils in the different LAs: 36% of pupils only rank one school, and other pupils ranked between 2 and 6 schools.

[3] There are additional allowances for children in care and children with one parent in the armed forces

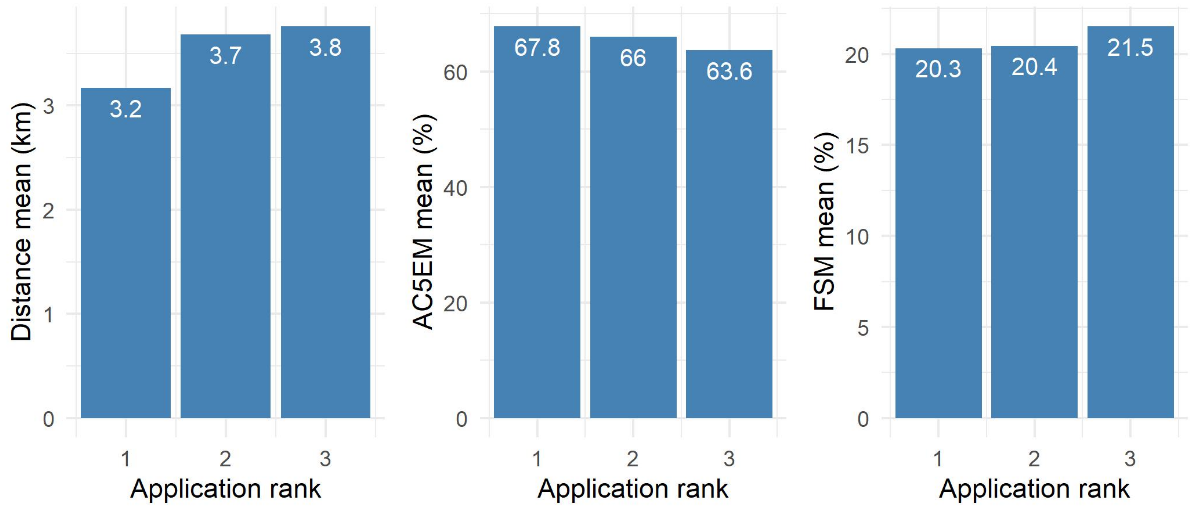
Table 1: **Number of applications submitted in the West Midlands in 2013 broken down by LA and length of lists**

	Length of list submitted					
Local authority	**1**	**2**	**3**	**4**	**5**	**6**
Birmingham	2,392	1,830	2,513	1,395	904	4,346
Coventry	918	890	1,780	SUPP	SUPP	SUPP
Dudley	728	1,299	744	194	68	97
Herefordshire	1,009	307	289	SUPP	SUPP	SUPP
Sandwell	1,401	853	787	281	144	239
Shropshire	1,705	558	308	SUPP	SUPP	SUPP
Solihull	781	544	440	159	215	SUPP
Staffordshire	3,744	1,774	1,343	10	10	SUPP
Stoke-on-Trent	1,040	648	917	SUPP	SUPP	SUPP
Telford and Wrekin	359	453	458	525	SUPP	SUPP
Walsall	1,000	906	609	233	331	SUPP
Warwickshire	2,475	1,342	811	251	75	47
Wolverhampton	831	647	569	262	396	SUPP
Worcestershire	1,574	851	822	SUPP	SUPP	SUPP
Total	**19,957**	**12,902**	**12,390**	**3,313**	**2,147**	**4,736**

Sources: author's calculations using ONS data (NPD application data for academic year 2014/2015). Notes: total number of applications is 55,445; entries marked 'SUPP' were suppressed to prevent recovery of counts under 10.

Figure 1 below shows the mean characteristics of schools ranked in the first three slots of applications submitted by households in the West Midlands in 2014. This shows that households tend to assign higher ranks to schools located closer, performing better, and with lower shares of disadvantaged pupils.

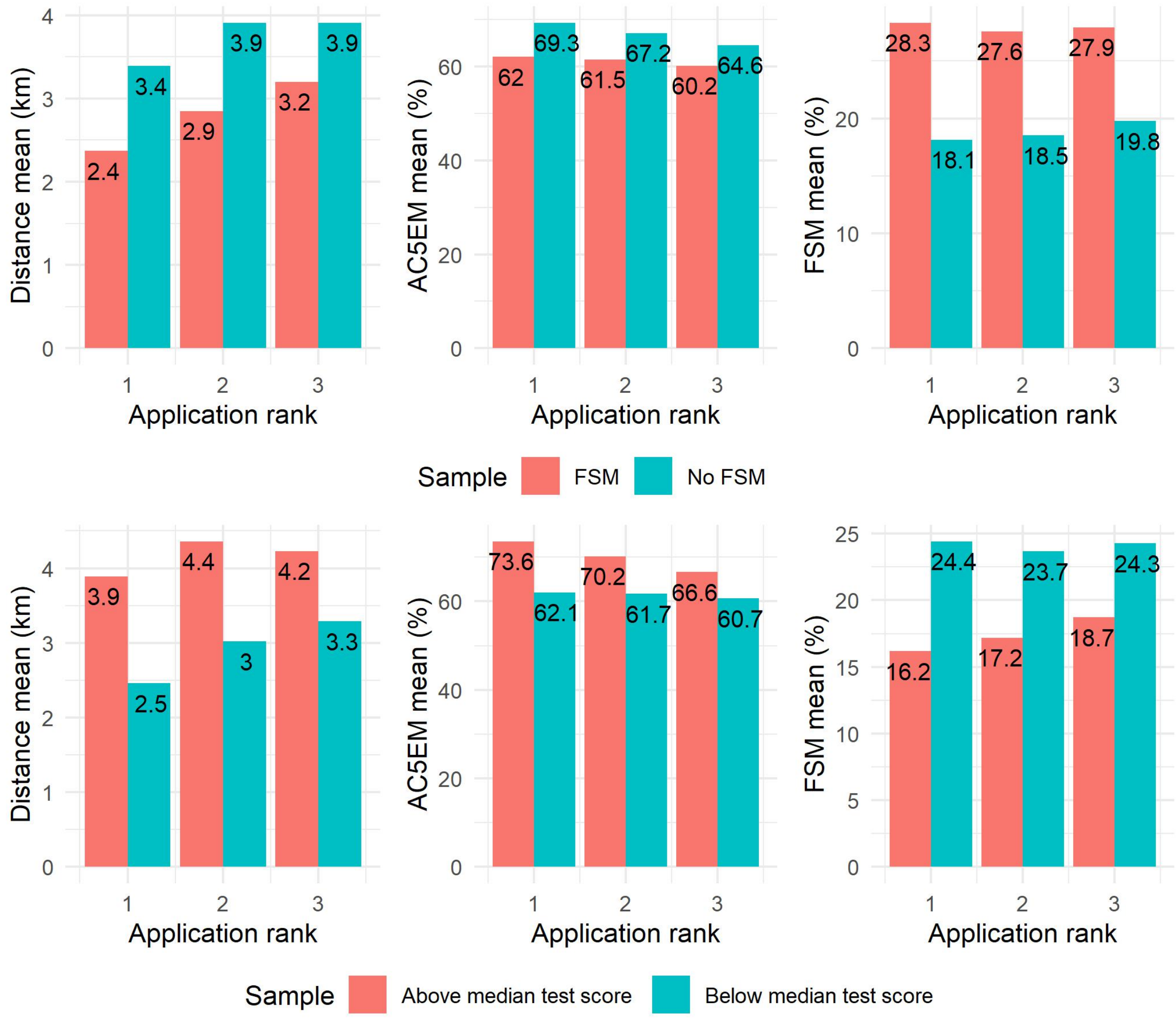
Figure 1: **Mean characteristics of schools ranked at slots 1 to 3 in applications**



Source: author's calculations based on school applications submitted in the West Midlands region in 2013 (all households ranking at least 3 schools). Notes: the number of households in this subsample is 21,713.

Figure 2 below shows patterns of heterogeneity in applications. Compared to applications submitted by affluent households, application submitted by disadvantaged households exhibit a steeper gradient in the distance variable (meaning that the average distance to listed schools increases more for less preferred slots), a flatter gradient in the performance variable (meaning that the average performance of listed schools increases less for less preferred slots), and a flatter gradient in the share of disadvantage pupils. The patterns are similar for households whose child had a primary school test score below the median, compared to other households. These patterns are consistent with the proposition that poorer households and households with lower educational attainment value school performance to a lesser degree, and similarly care about school composition to a lesser degree.

Figure 2: **Mean characteristics of schools ranked at slots 1 to 3 in applications, by type of household**



Source: author's calculations based on school applications submitted in the West Midlands region in 2013 (all households ranking at least 3 schools). Notes: the number of observations in this subsample were: 4625 households eligible for FSM, 17088 households ot eligible for FSMs, 10796 households above the median test score, 10917 households below the median test score.

3.2 School workforces

In 2019, there were 23,263 teachers in post in secondary schools in the West Midlands, 20,281 of whom were categorized as 'classroom teachers' (the others were occupying various senior leadership roles). Table 2 below shows some characteristics of the workforce for different quintiles of schools categorised by the share of disadvantaged pupils on their roll. This shows that the pupils-to-teacher ratio does not vary much across different categories of schools, but schools with more disadvantaged pupils tend to have higher shares of inexperienced and non-qualified teachers.

Table 2: **Pupils-to-teacher ratio (2019)**

School quantile	Average share of pupils eligible for FSMs (%)	Average pupils-to-teacher ratio	Average share of inexperienced teachers (%)	Average share of non-qualified teachers (%)
1	11.5	17.3	7.8	3.8
2	20.5	17.2	9.2	3.6
3	29.6	17	12.3	5
4	40.7	17.3	15.3	7.4
5	56.4	17.3	15.4	10.6

Source: author's calculations using SWFC data provided by ONS. Notes: sample is all teachers employed in the West Midlands region in 2019 with a permanent contract and not in a senior leadership role at that date (n=20281); inexperienced teachers are teachers with 3 years of experience or less.

Table 3 below shows that schools with a higher share of disadvantaged pupils tend to also experience much high turnover rates: the share of teachers employed in 2018 in the West Midlands who left their school of employment the following year was 13.3% for the lowest (least disadvantaged) quintile and 22.4% for the highest (most disadvantaged) quintile. Roughly one third of teachers leaving their jobs in 2018 moved to another school in the West Midlands, and the rest either left the state-funded education system, or moved to a school outside the West Midlands (a set of transition types that I will term the 'outside option' in the rest of this analysis).

Table 3: **Teacher turnover (2018-2019)**

School quantile	Average share of pupils eligible for FSMs (%)	Total number of quits	Turnover rate (%)	Quits to outside option (% of quits)	Quits to other school (% of quits)
1	11.1	539	13.3	67.0	33.0
2	19.7	582	15.2	65.3	34.7
3	28.4	550	15.7	64.0	36.0
4	38.7	784	19.5	56.5	43.5
5	55.7	1029	22.4	63.7	36.3

Source: author's calculations using SWFC data provided by ONS. Note: sample is all teachers employed in the West Midlands region in 2018 and not in a senior leadership role at that date (n=20006). A quit to the 'outside option' is defined as either the teacher leaving the database , or moving to a school outside the West Midlands region.

A higher turnover rate in more disadvantaged schools could merely reflect the fact that these schools tend to employ teachers who are at an earlier stage in their careers, and therefore plausibly more mobile. To investigate this possibility, I estimate a simple probit model of job quits as a function of observable teacher and school characteristics. Table 4 shows that, even controlling for teacher experience and qualification status, the probability of a job quit is higher when a teacher works in a more disadvantaged school.

Table 4: Probit model of exits as a function of school and teacher characteristics

	Model 1
Experience (log)	-0.068***
	(0.003)
Qualification status	-0.257***
	(0.025)
Senior management grade	-0.087***
	(0.010)
School FSM share	0.006***
	(0.000)
Urban school	-0.073***
	(0.007)
Num.Obs.	223569
BIC	209078.6
Log.Lik.	-104502.346

* p < 0.1, ** p < 0.05, *** p < 0.01

Estimation by maximum likelihood for teachers in the West Midlands 2010-2018

Table 5 provides more detail on the school-to-school transitions observed in the West Midlands in 2018. The table maps all the observed transitions by school quintile of origin (in rows) and school quintile of destination (in columns). For example the cell in the first row and second column indicates that 2.2% of the teachers who transitioned from one school to another in 2018 moved from a school in the first quintile (the least disadvanatged) to a school in the second quintile (the second least disadvantaged). In total, of the 1293 teachers who transitioned to another school in that year, 510 (40%) transitioned to a school in lower (less disadvantaged) quintile (corresponding to the cells highlighted in blue), and 388 (30%) transitioned to a school in a higher (more disadvantaged) quintile (corresponding to the cells highlighted in red).

Table 5: **School-to-school transitions (2018-2019)**

School quantile	**Transtions to quant 1**	**Transtions to quant 2**	**Transtions to quant 3**	**Transtions to quant 4**	**Transtions to quant 5**
1	4.7%	2.2%	3.4%	2.2%	1.2%
2	3.7%	4.2%	2.7%	2.9%	2.2%
3	3.9%	2.4%	2.7%	3.6%	2.8%
4	2.4%	3.4%	5.2%	8.5%	6.9%
5	3.2%	3.7%	5.6%	6.0%	10.4%

Source: author's calculations using DfE data (SWFC data for 2018-2019) provided by ONS. Note: Sample is all teachers employed in the West Midlands region in 2018 and not in a senior leadership role at that date (n=20006)

Table 6 shows the results of some simple regressions of individual wages on school and teacher characteristics. Model 2 shows that observationally equivalent teachers tend to be paid more in schools with higher shares of disadvantaged pupils. Compared to a teacher working in the least disadvantaged quintile, a teacher working in the most disadvantaged quintile earn approximately £1,500 more per year. This patterns holds when controlling for LA fixed effects (Model 3).

Table 6: **Regression of individual wages on school and teacher characteristics**

height	Model 1	Model 2	Model 3
Share of FSM pupils in school (log)	-0.034***	0.018***	0.019***
	(0.003)	(0.002)	(0.002)
Experience of teacher (log)		0.193***	0.193***
		(0.001)	(0.001)
Qualification status of teacher		0.217***	0.215***
		(0.007)	(0.007)
Num.Obs.	20112	20112	20112
R2 Adj.	0.007	0.613	0.617
LA fixed effect	No	No	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Sample is all classroom teachers employed in the West Midlandsregion in 2019. The dependent variable in all models is the full-time equivalent wage expressed in thousands GBP (logged). Standard errors are HC2

In summary, schools with larger shares of disadvantaged pupils tend to have higher shares of inexperienced teachers and higher turnover rates, despite offering higher wages conditional on teacher observable characteristics. When teachers move to different schools, they are more likely to move to schools with lower proportions of disadvantaged pupils. To rationalize all these observations, I use a model of job search where teachers have preferences over the composition of schools' pupil bodies in addition to wage. Teachers receive offers from schools at random rates, and move jobs when the value of an offer exceeds the value of job they hold (see section **??**).

4 Model of teachers and pupils sorting

An economy is populated by a set of pupils $\mathcal{P}$ indexed by i, a set of teachers $\mathcal{T}$ indexed by k, and a set of schools $\mathcal{S}$ indexed by j. To describe the model, I start by specifying the attributes of these different agents and their optimisation problems, before defining an equilibrium.

4.1 Pupils

I assume that each pupil and her household act as a unitary actor, and I refer to pupils and households interchangeably. Each pupil i is characterised by five variables: her residential location; her average test scores at the end of primary school, which I denote x_i^t; a binary variable indicating whether she is eligible for FSMs, which I denote x_i^f; and two unobserved taste shocks for school performance and composition, which I denote ν_i^p and ν_i^f, respectively. The residential location of households is determined exogenously.

Preferences

The utility that pupil i derives from attending school j is a linear function of five variables: the distance between the residential location of the pupil and the school, which is denoted d_{ij}; the school's AC5EM performance score, which is denoted z_j^p; the proportion of the school's pupil body eligible for FSMs, which is denoted z_j^f; an additional quality component denoted ξ_j which is observable to households but not to the econometrician; and a preference shock denoted ϵ_{ij}^u. The variable ξ_j could represent various attributes of the school that are valued by households, such as its ethos or the amount of bullying that takes place. The composition of a school's pupil body and its AC5EM score are endogenous variables, and the expectations of these variables are denoted $\hat{z}_j^f$ and $\hat{z}_j^p$, respectively. The utility function is:

$$u_{ij} = \beta_i^d d_{ij} + \beta_i^f \hat{z}_j^f + \beta_i^p \hat{z}_j^p + \xi_j + \epsilon_{ij}^u \tag{1}$$

where:

$$\begin{aligned}
\beta_i^d &= \beta^d + \beta^{df} x_i^f + \beta^{dt} x_i^t \\
\beta_i^f &= \beta^f + \beta^{ff} x_i^f + \beta^{ft} x_i^t + \nu_i^f \\
\beta_i^p &= \beta^p + \beta^{pf} x_i^f + \beta^{pt} x_i^t + \nu_i^p
\end{aligned}$$

$$(\nu_i^f, \nu_i^p) \sim \mathcal{N}(0, \Sigma)$$

and ϵ_{ij}^u is assumed to be independently distributed over pupils and schools with a type-1 extreme value distribution. This is a mixed multinomial logit model where preferences over school characteristics are allowed to vary between pupils based on their observable characteristics and, in the case of preferences over school performance and composition, unobservable taste variations. State schools do not charge fees in England, and this specification essentially quantifies preferences for school attributes in terms of willingness to travel for different categories of students.

When applying for a place in secondary school, each household ranks a subset of all schools in descending order of utility. This is a 'truth-telling' assumption, which I motivate in section 5.1 below. The subset of schools considered by a household is observed - it is the set of schools listed on its application - and fixed. The rank-ordered list of schools submitted by pupil i is denoted L_i.

Educational attainment

I model pupil outcomes (test scores) using a value-added framework of the type commonly used in the literature (most recently by Allende 2019 and Abdulkadiroglu et al. 2020). Let y_{ij}^c be the test score that pupil i would achieve at the end of secondary school if she attended school j.[4] This is determined by four variables: the pupil's test score at the end of primary school

[4] In practice, pupils obtain a test score for each subject they sit for their GCSE. These test scores are then aggregated by the DfE into two summary variables for each pupil: a continuous variable called Attainment 8, which is a weighted average of subject scores and which I denote y_{ij}^c; and a binary variable called AC5EM, which describes whether the pupil has attained grades of A to C in at least 5 subjects including English and Maths,

(denoted x_i^t as above); her FSM status (denoted x_i^f as above); a school quality component common to all pupils who attend the school, which is denoted θ_j; and an error term which is denoted ϵ_i^y. The test score production function takes the form:

$$y_{ij}^c = \gamma^t x_i^t + \gamma^f x_i^f + \mathbf{s}_i' \boldsymbol{\theta} + \epsilon_i^y \tag{2}$$

where $\mathbf{s}_i$ is a vector whose jth entry equals one if pupil i attends school j and zero otherwise, and $\boldsymbol{\theta}$ is a vector containing the value-added of the different schools. The child's FSM status can be interpreted as a proxy for unobserved contemporaneous inputs provided by the family (eg help with homework, private tutoring, etc) (Todd and Wolpin 2003).

4.2 Teachers

Teachers are likely to be heterogenous in terms of their preferences over school characteristics and their effectiveness (in terms of raising pupils' attainment) in a way that is likely to be only imperfectly correlated with teachers' observable characteristics. I use a discrete representation of this unobserved heterogeneity, based on Bonhomme, Lamadon, and Manresa 2022, whereby each teacher can belong to one of a finite number of 'categories', and all teachers in a given category share the same preferences and effectiveness up to some idiosyncratic shocks (in a sense that I will define more precisely below). Hence each teacher is characterised by her experience, denoted e, and her category, denoted c. There are C categories and there are T_c teachers in category c, and this is determined exogenously. A teacher's category is observable by schools but not by the econometrician.

Wage structure and preferences

The wage received by a teacher consists of two components: a 'regulated wage', denoted $w^r(e)$, which is set centrally by the Department for Education and represents the default wage payable to all teachers of experience e irrespective of their category or school of employment; and a 'market premium', denoted p_{cj}, which is set by schools and applies to all teachers of category c in school j (this can be positive or negative). That is, the wage of a teacher of category c and experience e in school j is:

$$w_{cje} = w^r(e) + p_{jc} \tag{3}$$

The flow utility received by teacher k of category c and experience e working for school j in period t is:

$$v_{jcekt} = \alpha_c^w w_{cje} + \alpha_c^f z_j^f + \zeta_{cj} + \epsilon_{kjt}^v \tag{4}$$

where z_j^f is the share of pupils eligible for FSMs at school j, ζ_{cj} is the utility value of the unobservable attributes of school j (this could represent the amenity value of the leadership's management style, the level of discipline enforced in the school, etc), and ϵ_{kjt}^v is a preference shock that is iid EV1 over teachers, schools, and years. That is, teachers care about wages, the

and which I denote y_{ij}^b. I use the continuous variable for the purpose of modelling pupil outcomes, and I assume that the binary variable is a function of the continuous variable and a normally-distributed random shock.

share of disadvantaged pupils enrolled in schools, and some unobservable characteristics. The preference parameters $(\alpha_c^w, \alpha_c^f, \zeta_{cj})$ are common to all teachers belonging to category c.

Given my specification for wages in (3), this can be reformulated as:

$$v_{jcekt} = \alpha_{ce} + \nu_{cj} + \epsilon_{kjt}^{v} \tag{5}$$

where $\alpha_{ce} = \alpha_c^w w^r(e)$ is a term that varies with experience but not over schools, and $\nu_{cj} = \alpha_c^w p_{jc} + \alpha_c^f \hat{z}_j + \zeta_{cj}$ is a term that varies over schools but not with experience.

A teacher can also work for the 'outside option', in which case her flow utility is $v_{c0e} + \epsilon_{k0t}$. Working for the outside option might involve working for a private school, working for a state school outside the region, or leaving the education sector entirely. I further assume that the flow value of the outside option increases with experience at the same rate as the flow value of schools, ie $v_{c0e} = \alpha_{ce} + \alpha_{c0}$. This is plausible given that for a significant share of teachers, the outside option involves working for another state school outside the area of interest or working for a private school (who would compete with state schools in the hiring of teachers).

Job search and value functions

To represent the process of matching teachers to schools I use a model of job search with posted wages, exogenous search effort, and random, on-the-job search, in the spirit of Burdett and Mortensen 1998. Teachers employed by a school sample job offers at rate λ_{c1}, while teachers working in the outside sector sample job offers at rate λ_{c0}. A teacher receiving an offer draws that offer from distribution f_c, where f_{cj} is the probability of receiving an offer from school j. Wages are not negotiated. There is an exogenous job destruction rate δ_c. Following Sorkin 2018, I assume that when teachers do not receive an offer and are not subject to a job destruction shock, they can decide whether to stay in their current role or move to the outside option.

The cross-sectional distribution of teachers of category c is g_c, where g_{cj} denotes the share of teachers working for school j. The share of teachers working for the outside option is r_{c0}. These quantities are determined endogenously.

Let $V_{cj} + \epsilon_{kjt}$ denote the present value of working for school j in period t for teacher k of category c, net of experience-related payoffs.[5] The model implies that V_{cj} takes the following form (in the remainder of this section I suppress the subscripts c):

$$\begin{aligned} V_j = \nu_j + \beta \Big\{ & \delta \int_{\epsilon_1} [V_0 + \epsilon_1] d\epsilon_1 \\ & + (1-\delta)\lambda_1 \sum_{s \in \mathcal{S}} f_s \int_{\epsilon_2} \int_{\epsilon_3} max[V_j + \epsilon_2, V_s + \epsilon_3] d\epsilon_2 d\epsilon_3 + \\ & + (1-\delta)(1-\lambda_1) \int_{\epsilon_4} \int_{\epsilon_5} max[V_j + \epsilon_5, V_0 + \epsilon_6] d\epsilon_5 d\epsilon_6 \Big\} \end{aligned} \tag{6}$$

That is, the present value of working for a school is the sum of the flow value of working at that school as defined in equation (5) and a continuation value. This continuation value is the probability-weighted sum of the value of the three possible events that can occur after the

[5] Given that the experience-related payoff is invariant over all employment options and enters the utility function additively, experience can essentially be ignored for the purpose of modelling teachers' decisions. Experience still affects teachers' allocation to school through the search process - ie less experienced teachers are more likely to work for less desirable schools because they have ot received an offer from more desirable schools yet.

current period: at rate δ, the teacher is affected by an exogenous job-destruction rate, in which case she obtains the expected value of the outside option; at rate $(1-\delta)\lambda_1$, the teacher receives an offer from another school, in which case she obtains the expectation of the maximum of the value of her current job and the value of the offered job; and at rate $(1-\delta)(1-\lambda_1)$ she makes a utility maximizing choice between staying in her current job and moving to the outside option. Given that the preference shocks are iid EV1, this can be expressed more simply as:

$$\begin{aligned} V_j = \nu_j + \beta \Big\{ & \delta[V_0 + E] \\ & + (1-\delta)\lambda_1 \sum_{s \in \mathcal{S}} f_s [log(e^{V_j} + e^{V_s}) + E] \\ & + (1-\delta)(1-\lambda_1)[log(e^{V_j} + e^{V_0}) + E] \Big\} \end{aligned} \tag{7}$$

where E is Euler's constant.

Similarly, the forward-looking value of the outside option is defined as follows:

$$\begin{aligned} V_0 = \alpha_0 + \beta \Big\{ & \lambda_0 \sum_{s \in \mathcal{S}} f_s \int_{\epsilon_6} \int_{\epsilon_7} max[V_0 + \epsilon_6, V_s + \epsilon_7] d\epsilon_6 d\epsilon_7 + \\ & + (1-\lambda_0) \int_{\epsilon_8} [V_0 + \epsilon_8] d\epsilon_8 \Big\} \end{aligned} \tag{8}$$

This form reflects the fact that, when employed in the outside sector, a teacher can either receive an offer from a school in the inside set, or stay in the outside sector.

Labour supply function

The structure of the search model together with the forward-looking value of schools induces the labour supply functions applicable to each school (each school faces one labour supply function in each category). In the steady state, the number of teachers leaving school j in any period must equal the number of teachers joining that school:

$$\begin{aligned} & T(1-r_0)g_j \Big[\delta + (1-\delta)\lambda_1 \sum_{s \in \mathcal{S} \backslash j} f_s \frac{e^{V_s}}{e^{V_s} + e^{V_j}} + (1-\delta)(1-\lambda_1) \frac{e^{V_0}}{e^{V_j} + e^{V_0}} \Big] \\ & = T(1-r_0) \sum_{s \in \mathcal{S} \backslash j} g_s (1-\delta)\lambda_1 f_j \frac{e^{V_j}}{e^{V_s} + e^{V_j}} + T r_0 \lambda_0 f_j \frac{e^{V_j}}{e^{V_j} + e^{V_0}} \end{aligned} \tag{9}$$

Dividing by $T(1-r_0)(1-\delta)$ throughout and rearranging gives the share of teachers working in school j in the steady state:

$$g_j = \frac{f_j \left(\lambda_1 \sum_{s \in \mathcal{S} \backslash j} g_s \frac{e^{V_j}}{e^{V_s} + e^{V_j}} + \frac{r_0}{1-r_0} \frac{1}{1-\delta} \lambda_0 \frac{e^{V_j}}{e^{V_j} + e^{V_0}} \right)}{\frac{\delta}{1-\delta} + \lambda_1 \sum_{s \in \mathcal{S} \backslash j} f_s \frac{e^{V_s}}{e^{V_s} + e^{V_j}} + (1-\lambda_1) \frac{e^{V_0}}{e^{V_j} + e^{V_0}}} \tag{10}$$

This depends on the share of teachers working for the outside option, r_0. This is found in the same way, by setting out the flow-balance equation for teachers leaving and entering the outside sector, and rearranging:

$$
\begin{aligned}
Kr_0\lambda_0\sum_{s\in\mathcal{S}} f_s\frac{e^{V_s}}{e^{V_s}+e^{V_0}} &= K(1-r_0)\Big[\delta+(1-\delta)(1-\lambda_1)\sum_{s\in\mathcal{S}} g_s\frac{e^{V_0}}{e^{V_s}+e^{V_0}}\Big] \\
\Rightarrow r_0 &= \frac{\delta+(1-\delta)(1-\lambda_1)\sum_{s\in\mathcal{S}} g_s\frac{e^{V_0}}{e^{V_s}+e^{V_0}}}{\delta+(1-\delta)(1-\lambda_1)\sum_{s\in\mathcal{S}} g_s\frac{e^{V_0}}{e^{V_s}+e^{V_0}}+\lambda_0\sum_{s\in\mathcal{S}} f_s\frac{e^{V_s}}{e^{V_s}+e^{V_0}}}
\end{aligned} \tag{11}
$$

Therefore, the total supply of teachers of a particular category to school j is simply $l_j = T(1-r_0)g_j$.[6] Overall, for each category, each school faces a labour supply curve of the form:

$$
l_j(p_j) = l(p_j, \hat{\mathrm{p}}_{-j}, \hat{\mathrm{g}}, \hat{\mathrm{z}}, \mathrm{f}, \alpha_0, T, \lambda_1, \lambda_0, \delta) \tag{12}
$$

where $(\hat{\mathrm{p}}_{-j}, \hat{\mathrm{g}}, \hat{\mathrm{z}})$ denote the school's expectations over the premiums posted by other schools, the share of teachers hired by schools, and the share of disadvantaged pupils taught by schools.

4.3 Schools

Each school faces an 'education production function' where its value added θ_j depends on the number of teachers it employs in each category divided by the total number of pupils enrolled at the school (denoted q_j).[7] I assume that this production function has a Constant Elasticity of Substitution (CES) form:

$$
\exp(\theta_j) = \gamma_j^0\left(\sum_c \gamma_{jc}\Big(\frac{l_{jc}}{q_j}\Big)^{\sigma}\right)^{\frac{1}{\sigma}}\epsilon_j^{\theta} \tag{13}
$$

where γ_{cj} is the effectiveness of teachers in category c in school j. At this stage, this is left unrestricted, allowing for flexible 'matching effects' (whereby the effectiveness of each category of teacher can vary across schools in a flexible way).

Each school receives a budget allocation that is a function of the size and composition of its pupils body, which I denote $B(q_j, z_j^f)$. Schools flex the wages they post to the different categories of teachers to attract the optimal set of teachers given the composition of their pupil body. Formally, noting that maximizing θ_j is the same as maximizing the term in brackets in equation (13), their problem is:

$$
\max_{\mathrm{p}_j}\sum_c \gamma_{jc}\Big(\frac{l_{jc}}{q_j}\Big)^{\sigma} \quad st \quad \sum_c l_{jc}(p_{jc})w_{jc}(p_{jc}) \leq B_j \tag{14}
$$

where $w_{jc}(p_{jc}) = \sum_e h_{ce}\cdot(w^r(e)+p_{jc})$ is the experience-weighted cost of hiring a teacher of category c.

Assuming an interior solution, the first order condition with respect to p_{jc} is:

$$
\gamma_{jc}\sigma l_{jc}^{\sigma-1}\frac{\partial l_{jc}(p_{jc})}{\partial p_{jc}} - \lambda_j\left[\frac{\partial l_{jc}(p_{jc})}{\partial p_{jc}}w_{jc}(p_{jc}) + l_{jc}(p_{jc})\right] = 0 \tag{15}
$$

We can find an expression for the Lagrange multiplier λ_j by multiplying each FOC by

[6] The number of teachers of experience e within that group is $l_{jce} = h_{ce}l_{jc}$ where h_{ce} is the proportion of teachers of experience e in category c in the economy.

[7] Section **??** explains how the value added affects pupil outcomes

$w_{jc}(p_{jc})$, summing them up over categories, and simplifying using the budget constraint. This gives:

$$\lambda_j = \frac{\sum_{c'} \gamma_{jc'} \sigma l_{jc'}^{\sigma-1} l'_{jc'} w_{jc'}(p_{jc'})}{\sum_{c'} l'_{jc'} w_{jc'}(p_{jc'})^2 + B_j} \tag{16}$$

Substituting this expression in (15) gives a system of $S \times C$ equations in $S \times C$ unknowns for the economy. The derivatives of labour supply with respect to price are the solutions to a system of linear equations obtained by totally differentiating the value functions.

Each school has a fixed capacity, and the capacities of the different schools are collected in vector $\mathbf{r}$. If a school is oversubscribed, it ranks all pupils belonging to the lowest coarse priority group by distance in ascending order (that is, it prioritises pupils who live closer). Local authorities allocate school places to pupils using a deferred-acceptance algorithm. The resulting allocation of pupils to schools is represented by the logical matrix $\mathbf{A}$, where element $A_{ij} = 1$ if pupil i is allocated to school j, and 0 otherwise. The allocation is a deterministic function of the rank-ordered lists submitted by pupils, their distance to schools, and schools' capacity:

$$\mathbf{A} = g(\{L_i\}_{i \in P}, \{d_{ij}\}_{i \in P, j \in S}, \mathbf{r}) \tag{17}$$

In turn, the allocation determines the size and composition of schools' pupil body and, together with the test score production function, their performance metrics.:

$$q_j = \sum_{i \in P} I[A_{ij} = 1] \quad z_j^f = \frac{1}{q_j} \sum_{i \in P^f} I[A_{ij} = 1] \quad z_j^p = \frac{1}{q_j} \sum_{i \in P} I[A_{ij} = 1] y_{ij}^b$$

4.4 Sorting equilibrium

An equilibrium is defined by three objects: the ordered lists submitted by households $\{L_i\}_{i \in \mathcal{P}}$; the wages set by schools for the different categories of teachers $\{\mathbf{w}_j\}_{j \in \mathcal{S}}$; and an allocation of teachers to schools $\{\mathbf{t}_j\}_{j \in \mathcal{S}}$. The tuple $\left\{\{L_i\}_{i \in \mathcal{P}}, \{\mathbf{w}_j\}_{j \in \mathcal{S}}, \{\mathbf{t}_j\}_{j \in \mathcal{S}}\right\}$ is a rational-expectations sorting equilibrium if the following conditions are satisfied:

- the ordered lists $\{L_i\}_{i \in P}$ reflect households' consideration sets and preferences as defined in (1);
- the wages set by schools $\{\mathbf{w}_j\}_{j \in S}$ solve their optimisation problem defined in (14);
- the allocation of teachers to school $\{\mathbf{t}_j\}_{j \in S}$ reflects the preferences given in (4); and
- expectations are rational, that is $\hat{q}_j = q_j$, $\hat{z}_j^f = z_j^f$, and $\hat{z}_j^p = z_j^p$.

5 Estimation

5.1 Household preferences

My main identifying assumption is that the rank-ordered lists submitted by pupils truthfully reflect their preferences. This assumption is motivated by the fact that the deferred acceptance mechanism used in England is strategy-proof, in the sense that truth-telling is a weakly dominant strategy if there are no application costs (Dubins and Freedman 2018). Fack, Grenet, and He 2019 show that truth-telling is the *unique* Bayesian Nash equilibrium under deferred acceptance if and only if there are no application costs and the joint distribution of preferences and priorities has full support (ie there is uncertainty in admission outcomes for all schools considered). These conditions might not hold for all pupils in all local contexts. In particular, they fail to hold for pupils who consider more schools than they can rank. To mitigate this risk, I focus on the region (outside London) where LAs allow pupils to rank the largest number of schools, which is the West Midlands. On average, pupils in the West Midlands can rank 5.2 schools, and only 18% of them rank the maximum number of schools allowed in their home LA.

Even with the truth-telling assumption, the estimation of preferences presents the usual challenge of endogeneity since observed school characteristics may be correlated with unobserved characteristics. Suppose for example that school leaders who use a more effective curriculum are also more effective at enforcing school discipline and reducing the amount of bullying. Then the process of school choice combined with the test score production function induces a positive correlation between performance scores z_j^p and unobservable characteristics ξ_j. Alternatively, some school leaders might seek to achieve high test scores by 'teaching to the test', offering a narrower curriculum, or focusing effort on marginal students, which could induce a negative correlation between performance scores and unobservable characteristics.

To address these endogeneity issues I proceed in two steps, in the spirit of the IO literature on differentiated product markets. In the first step I re-specify equation (1) with an alternative-specific constant for each school that subsumes the average utility of unobserved characteristics, the share of disadvantaged pupils and the performance score. This is denoted:

$$\delta_j = \beta^f x_j^f + \beta^p x_j^p + \xi_j \tag{18}$$

I can estimate these alternative-specific constants and the remaining parameters in the utility function using maximum simulated likelihood. Only differences in utility matter in this model, so I normalise the alternative-specific constant of one of the schools to zero.

In the second step I use these estimates of δ_j together with instrumental variables for the variables in equation (18) to estimate the coefficients β^f and β^p and the residual for each school ξ_j. My instruments for these two endogenous variables are based on the location of households and schools, ie local demographics and local market structures. These variables will evidently have an effect on the share of disadvantaged pupils enrolled in schools: a school located in a neighbourhood with more disadvantaged households will enroll a higher proportion of disadvantaged pupils. But these variables should also have an effect on the measured performance of schools: if the attainment of a child in secondary school depends on her socio-economic status and prior attainment, as posited by my model of educational attainment (2), then a school located close to large numbers of children from affluent families and with high prior attainment - and further away from other schools which might also attract such children - will achieve higher performance scores, independently of its unobservable characteristics. The identifying assumption here is that the location of households relative to schools is independent of unobservable school quality conditional on observed household demographics, that is $\xi_j \perp d_{ij} | x_i^t, x_i^f$. This approach is somewhat analogous to the logic of 'BLP instruments' (after Berry, Levinsohn, and Pakes 1995).

I use approximations to optimal instruments to improve the efficiency of the estimation procedure. If $\boldsymbol{Z}$ collects the instruments (the location and charateristics of households and schools), θ collects the parameters in (18), and θ^0 is the true value of these parameters, then the optimal IVs are given by (Chamberlain 1987):

$$z_{nj}^* = E\left[\frac{\partial \xi_j(\theta^0)}{\partial \theta}\middle| \boldsymbol{Z}\right] \quad n = 1, ..., dim(\theta) \tag{19}$$

These optimal instruments are not feasible since the true value of the parameters and the distribution of the unobserved term are unknown, but they can be approximated heuristically. I use the approximation proposed by Reynaert and Verboven 2014, which sets $\xi = 0$ and uses a guess of parameter values based on simpler estimation methods. Concretely, I use OLS to provide initial guesses of the parameters in (18) and (2) (with the exception of school value added, as it is endogenous), and I simulate the choices of households and the resulting characteristics of schools iteratively until I find a fixed point in the value of school characteristics. One can think of these instruments as the component of school performance and composition that purely reflects local geographies and demographics.

Table 7 below shows estimates of the parameters estimated by simulated maximum likelihood (in the first step of the estimation procedure). This shows that different types of households make different trade offs when applying to schools: disadvantaged households put more weight on distance (ie face higher travel costs), less weight on school performance, and less weight on the share of disadvantaged pupils; households with higher primary test scores put less weight on distance, more weight on school performance, and more weight on the share of disadvantaged pupils.

Table 7: **Estimates of demand model parameters obtained by maximum simulated likelihood**

	Parameter	Coeff.	Std.Err
Panel A: observed heterogeneity			
Distance	β^{d}	-0.191	0.003
Distance x pupil FSM status	β^{df}	-0.055	0.006
Distance x pupil primary test score	β^{dt}	0.047	0.002
School AC5EM x pupil FSM status	β^{pf}	-0.894	0.149
School AC5EM x pupil primary test score	β^{pt}	1.256	0.064
School FSM share x pupil FSM status	β^{ff}	0.505	0.144
School FSM share x pupil primary test score	β^{ft}	-0.189	0.065
Panel B: unobserved heterogeneity			
School AC5EM - variance	σ^{p}	4.07	0.894
School FSM share - variance	σ^{f}	2.44	0.708
Covariance	σ^{pf}	-1.271	0.586

Sources: parameters estimated by maximum simulated likelihood, number of observations 77250

Table 8 below shows estimates of the demand model parameters estimated by IV regression (the second step of the estimation procedure). The coefficient on school performance is higher when estimated using IV than when estimated using OLS. This suggests that there is a negative correlation between school performance and unobserved school characteristics. The coefficient on school composition becomes statistically insignificant. The results of the first stage regression are provided in table 9 below.

Table 8: **Estimates of demand model parameters obtained by IV**

		OLS	IV
(Intercept)		−1.037***	−1.843***
		(0.228)	(0.584)
School AC5EM	β^p	1.929***	3.108***
		(0.325)	(0.820)
School FSM	β^f	−0.965**	−0.425
		(0.260)	(0.535)
Num.Obs.		342	342
R2		0.467	0.444
se_type		HC2	HC2

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 9: **First stage regression for IV model used in household preferences**

	School AC5EM score	School FSM share
(Intercept)	0.017	0.035
	(0.063)	(0.048)
School AC5EM score - expected	1.011***	−0.034
	(0.092)	(0.068)
School FSM share - expected	0.148**	0.829***
	(0.072)	(0.063)
Num.Obs.	342	342
R2	0.467	0.710
DF Resid	339	339

* p < 0.1, ** p < 0.05, *** p < 0.01

5.2 School value added

A common approach to estimating school value-added is to adopt a 'selection on observables' restriction. Considering equation (2), this is $E[\epsilon_i^y | y_i^p, x_i^f, \mathbf{s}_i] = 0$. This assumption implies that an Ordinary Least Square (OLS) regression of individual test scores on school indicators and individual characteristics recovers consistent estimates of γ^p, γ^f and $\boldsymbol{\theta}$.

In our context, selection on observables is a strong assumption. The error term ϵ_i^y is likely to reflect the effect of a range of inputs into a child's education that are not captured by her FSM status and primary school test score, for example: (i) past educational inputs from the family or schools; (ii) contemporaneous inputs from the family; and (iii) the child's innate ability.[8]

[8]Todd and Wolpin 2003 provide a framework for assessing the validity of value-added models. Suppose that the true technology is $y_{ia} = X_{ia}\alpha_1 + X_{ia-1}\alpha_2 + ... + X_{i1}\alpha_a + \beta\mu_{i0} + \epsilon_{ia}$ where y_{ia} is the achievment of child i at age a, X_{ia} is the vector of educational inputs applied at age a, and μ_{i0} is the child's endowment at birth. Subtracting γy_{ia-1} from both sides and collecting terms gives: $y_{ia} - \gamma y_{ia-1} = X_{ia}\alpha_1 + X_{ia-1}(\alpha_2 - \gamma\alpha_1) + ... +$

I refer to these different factors as a child's unobserved 'ability'. A particular concern in my context is that the process of school choice may induce a correlation between $\mathbf{s}_i$ and ϵ_i^y, in the sense that the pupils enrolled at different schools might have different average abilities. This might be the case for example if the households who value attainment more tend to apply to schools with higher value added *and* provide more education inputs outside school (eg private tutoring). In that scenario, estimating equation (2) using OLS would overstate differences in value added between schools.[9]

A second approach is to use the information contained in pupils' applications to correct for the selection bias. Recall that the demand model specified in equation (1) allows for unobserved variations in preferences for performance, captured in ν_i. Suppose that a pupil's unobserved ability is correlated with the intensity of her preference for performance, which is what may induce the selection bias in the first place. That is, we can decompose the error term in equation (2) as $\epsilon_i^y = \tau^\nu \nu_i + \tilde{\varepsilon}_i^y$, implying:

$$y_{ij}^s = \gamma^p y_i^p + \gamma^f x_i^f + \mathbf{s}_i' \boldsymbol{\theta} + \tau^\nu \nu_i + \tilde{\varepsilon}_i^y \tag{20}$$

The variable ν_i is not observed, but the mixed logit model identifies the distribution of that coefficient for each individual conditional on her observed school choices L_i. This is (eg Train 2009):

$$h(\nu|L_i, \sigma, \mu) = \frac{P(L_i|\nu)\phi(\nu|\sigma, \mu)}{\int P(L_i|\nu)\phi(\nu|\sigma, \mu)d\nu} \tag{21}$$

where $P(L_i|\nu)$ is the probability of individual i submitting the rank-ordered list L_i for a given value of ν, and $\phi(\nu|\sigma, \mu)$ is the distribution of ν in the population. The expectation of a pupil's preference for performance conditional on her rank-ordered list L_i is then:

$$\overline{\nu}_i = \int \nu \cdot h(\nu|L_i, \sigma, \mu)d\nu \tag{22}$$

If we are willing to assume that $E[\tilde{\varepsilon}_i^y | y_i^p, x_i^f, \mathbf{s}_i, L_i, \sigma, \mu] = 0$, which is a weaker assumption than selection on observables, then we can replace ν_i by $\overline{\nu}_i$ in equation (20) and estimate the parameters using OLS. This is essentially a control function approach. The intuition for how this can identify the value-added of schools is simple: if a household ranks better-performing schools higher than can be predicted by this household's observed characteristics, then one can infer that this household has a high unobserved preference for school performance, and this information can be used to control for this household's unobserved ability.

One limitation of this approach in my context is that I can only compute $\overline{\nu}_i$ for households who rank at least two schools in their application (because equation (21) is only defined for a non-trivial choice set of at least two schools). It is conceivable that the number of schools ranked by a household is correlated with the unobserved ability of their children, for example if

$X_{i1}(\alpha_a - \gamma\alpha_{a-1}) + (\beta_a - \gamma\beta_{a-1})\mu_{i0} + \epsilon_{ia} - \gamma\epsilon_{ia-1}$. This boils down to the value-added model in equation (2) only if the coefficients associated with past inputs and the endowment decline at the same rate with distance from the date of measurement, ie if $\alpha_a - \gamma\alpha_{a-1} = 0$ and $\beta_a - \gamma\beta_{a-1} = 0$

[9] Alternatively, parents may see school and family inputs as substitutes rather then complements (eg they may decide to provide *less* tutoring if the child attends a high-value added school). In that scenario, estimating equation (2) using OLS would *understate* differences in value added between schools

households who value educational attainment more also research more schools and provide more tutoring at home. If that is the case, the control function approach risks introducing a sample selection issue. To address this issue I use a standard sample selection model (Heckman 1979), where a household's propensity to rank more than one school is a function of her observable characteristics (whether she is eligible for FSMs and her primary school test scores) and the number of schools present within a radius of 2km.

I estimate four models: one model that only controls for pupils' test scores in primary school (a simple 'selection on observables' model);[10] a second model that also controls for pupils' FSM status (a slightly richer 'selection on observables' model), a third model that also controls for pupils' estimated mean unobserved preference for school performance (a 'control function' model); and a fourth model that also control for factors affecting a household's propensity to rank more than one school (a control function and sample selection model). The coefficients on pupil characteristics are reported in Table 10. The coefficients on control factors are both positive and statistically significant, indicating that a household's unobserved 'ability' is indeed positively correlated with her unobserved preference for school performance and her propensity to list more than one school.

These results also show that school value added is an important determinant for educational attainment: moving a pupil from a school at the 25th percentile of the value added distribution to a school at the 75th percentile improves her test scores by 0.33 standard deviation. By way of comparison, moving a pupil out of FSM status increases her test scores by 0.22 standard deviation; and moving a pupil from the 25th to the 75th percentile of the distribution of primary school scores increases her secondary test scores by 0.79 standard deviation.

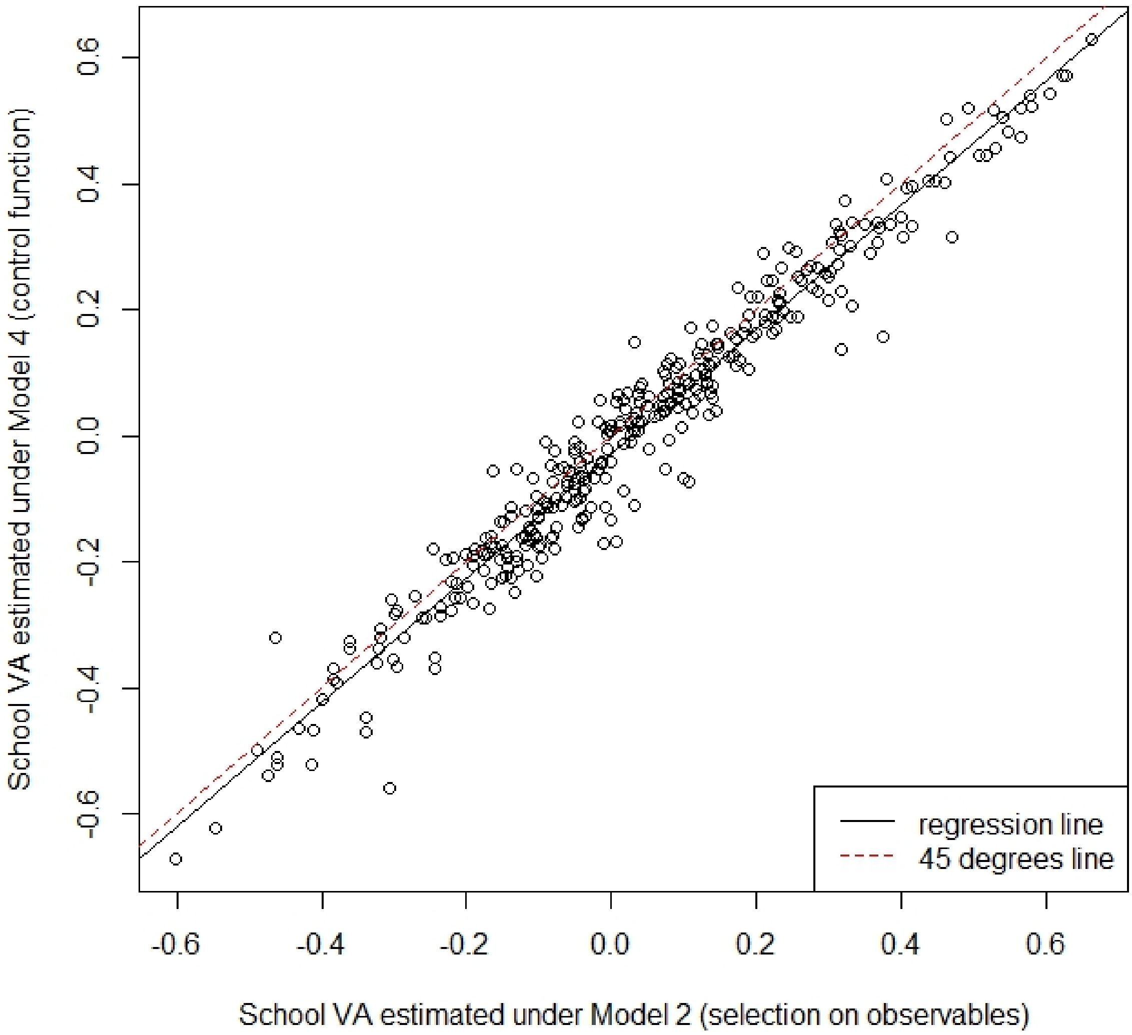
[10] this is the approach that underpins the DfE's Progress 8 measure of value added

Table 10: **Estimates of test score determinants**

	Model 1 selection on observ.	Model 2 selection on observ.	Model 3 control function	Model 4 control function
Pupil primary school score	0.688***	0.676***	0.682***	0.688***
	(0.003)	(0.003)	(0.004)	(0.005)
Pupil FSM status		-0.221***	-0.220***	-0.223***
		(0.008)	(0.010)	(0.009)
Pupil unobserved preference			0.007***	0.007***
			(0.002)	(0.002)
Variance (selection model)				0.624***
				(0.003)
Covariance (selection model)				0.112***
				(0.036)
Num.Obs.	48233	48233	31397	31397
R2	0.601	0.608	0.625	
se_type	HC2	HC2	HC2	
School dummy	Yes	Yes	Yes	Yes

Interest resides in how estimates of school value added that control for unobservable factors compare with more simple estimates that only control for observable factors (and which are more commonly used by policy-makers). Figure 3 below compares estimates of value added obtained under model 4 with estimates from model 1 (which most closely approximates the 'Progress 8' measure of value added introduced by the DfE in 2016). This shows that, while estimates of value-added differ between the two models for many schools, there is no obvious pattern whereby the simpler model significantly overstates differences in value added between schools (if this were the case, then the slope of the regression line would be significantly below 1).

Figure 3: **Estimates of school value added obtained under different assumptions**



For completeness, Table 11 below shows the results of a series of simple linear regressions of value added estimates obtained under models 2,3, and 4 on estimates obtained under models 1 and 2. The coefficients are only slightly below 1 and not statistically different from 1.

Table 11: **Comparison of different estimates of value added (linear regressions)**

	Model 2 estimates	Model 3 estimates	Model 4 estimates
(Intercept)	0.045	0.012	−0.027
	(0.001)	(0.003)	(0.003)
Model 1 estimates	0.978		
	(0.005)		
Model 2 estimates		0.980	0.986
		(0.012)	(0.012)
Num.Obs.	337	337	337
R2	0.989	0.956	0.954
R2 Adj.	0.989	0.956	0.954
$\text{se}_t ype$	HC2	HC2	HC2

Overall, this analysis suggests that, while the simple 'Progress 8' measure of value added is likely to underestimate the value added of schools with large shares of disadvantaged pupils, it does not otherwise suffer from selection biases that might overstate differences in value-added between schools.

5.3 Teacher categories

My estimation of teacher's preferences proceeds in three steps. Following the group fixed effect approach proposed by Bonhomme, Lamadon, and Manresa 2019, I start by recovering the unobserved category of each individual teacher using a clustering algorithm based on relevant moments of outcome variables. In a second stage, taking teachers' estimated categories as given, I estimate their valuation of working at different schools based on observed patterns of transitions between schools (and between schools and the outside option). Lastly, I project estimated school values on observed characteristics to recover the parameters of teachers' utility function. The remainder of this section provides more detail on these three steps.

The first step in my estimation procedure is to recover an estimate of the category of each teacher using a variant of the k-means algorithm. The k-means algorithm classifies individuals into C groups based on individual-specific moments m_k. That is, it assigns a category $\hat{c} \in 1, ..., C$ to each teacher to satisfy the following condition:

$$\big(\hat{c}(1), ..., \hat{c}(K)\big) = \underset{\{c(k)\}_{k=1:K}}{\text{argmin}} \sum_{k=1}^{K} \sum_{m=1}^{M} \left(h_k^m - \bar{h}^m(c(k))\right)^2$$

where $\bar{h}^m(c)$ is the mean of moment m in group c. This simply assigns categories to minimise the total distance between individual moments and group averages. The individual moments must be informative about the underlying types. I use two moments: the average wage premium received by a teacher over the period of observation;[11] and the average share of disadvantaged

[11] The wage premium received by a teacher in a given year is the difference between her full-time equivalent wage in that year and the regulated wage for her experience level. The regulated wage for a given experience level $w^r(e)$ is estimated as the average wage received by teachers of that level of experience before 2014 (the year

pupils at the schools where a teacher was employed over the period of observation. The idea is that, if teachers in different categories have different levels of effectiveness and preferences for school characteristics, and if schools observe preferences and effectiveness when setting wages, then teachers belonging to different categories are likely to work in different types of schools and earn different levels of wages. Conversely, teachers earning similar levels of wages in similar schools should belong to the same category.

One difficulty in my context is that I observe teachers at different stages in their careers, and their search activity implies that their wages and the characteristics of the schools they work for might vary over time in systematic patterns, even if their category is fixed. If teachers dislike working in environments with a lot of disadvantaged pupils, they will progressively move to schools with fewer disadvantaged pupils and/or higher wage premiums, even if their preferences and effectiveness remain fixed. In this context, the standard k-means algorithm will create clusters that capture a mix of individual heterogeneity and experience effects. To mitigate this issue, I follow Jolivet and Postel-Vinay 2020 and implement a modified version of k-means that essentially allows the expectations of the moments to vary with experience within each group. That is, I consider the following problem:

$$\underset{\{c(k)\}_{k=1:K},\{\rho_c^{0,h},\rho_c^{1,h},\rho_c^{2,h}\}_{c=1:C,h=1:H}}{\operatorname{argmin}} \sum_{k=1}^{K}\sum_{m=1}^{M}\left(h_k - (\rho_{c(k)}^{0,m} + \rho_{c(k)}^{1,m}\cdot e_k + \rho_{c(k)}^{2,m}\cdot e_k^2)\right)^2 \tag{23}$$

This algorithm is implemented in three steps. First, I set an initial partition (based on teachers' position in the distribution of wage premiums). Second, for each category and for each moment (wage premium and school FSM) I regress the value of the moment for individual teachers placed in that category on a constant, their experience, and their experience squared. Third, taking estimates of the ρ parameters as given, I update the partition. I iterate on steps 2 and 3 until the partition is stable.

My model assumes that a teacher's category is fixed over time. To make this assumption more plausible I use only the last three years of my panel in the estimation (2017-2019). Having estimated the category of each teacher, I estimate the wage premium posted by each school to each category as the median premium observed for teachers of that category in that school. Schools employ a small number of teachers (50 on average), so I fix the number of categories to 2 to limit the amount of noise in the estimation of the median wage.

Table 12 below shows the mean characteristics of teachers clustered in two categories using this procedure. Compared to teachers in the first category, teachers in the second category work for less disadvantaged schools and earn lower wage premiums, despite being slightly more experienced.

the wage reforms described in section 2 were introduced) in LA-maintained schools (the schools subject to the regulated wage guidance), controlling for LA fixed effects and qualification status

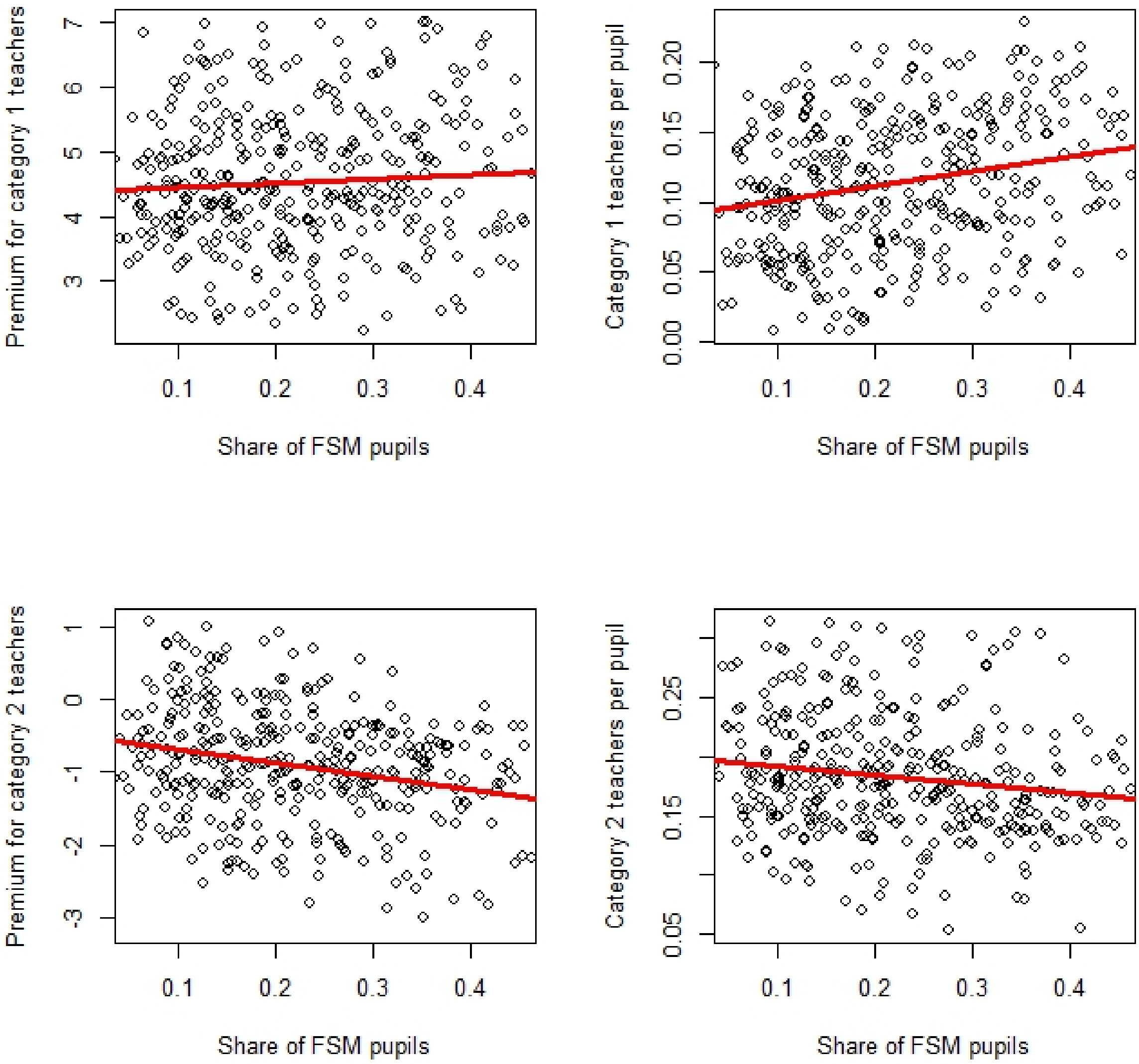
Table 12: **Mean characteristics of teachers clustered into two categories using modified k-means algorithm**

	Category 1			Category 2			
	N	Mean	SD	N	Mean	SD	Test
Wage premium	9165	4.052	3.147	16034	-1.888	2.731	F=24644.25***
School FSM share	9165	34.19	16.751	16034	31.432	16.462	F=161.597***
Teacher experience	9165	10.928	6.378	16034	9.422	7.064	F=284.124***
Ethnicity - White	8282	0.857	0.35	14113	0.837	0.369	F=15.764***
Holds degree in STEM	7153	0.37	0.483	12029	0.35	0.477	F=8.31***
Holds MSc or Phd	7153	0.1	0.301	12029	0.092	0.288	F=4.092**

Notes: sample is all teachers present in the West Midlands over 2017-2019

Figure 4 below shows how, within each category, wages and allocations vary across school types. Each dot on these chart is a school: the x-axis reports the FSM share of that school, and the y-axis reports the median premium paid at that school for teachers of a particular category (left panel), or the teachers-per-pupil ratio at that school (right panel). The two charts at the top relate to teachers in category 1, while the two charts at the bottom relate to teachers in category 2. Overall, the charts show that more disadvantaged schools tend to offer slightly higher premia to teachers of category 1, and slightly lower premia to teachers of category 2. They also tend to hire larger shares of category 1 teachers and lower shares of category 2 teachers. Within my model of job search, these patterns in wages and allocations can be explained by three factors: differences in preferences, differences in effectiveness, and differences in job offer rates. The following subsections seek to disentangle the relative importance of these factors.

Figure 4: **Distribution of wage premia and allocations across schools for the two categories of teachers**



5.4 School values

The second step in the estimation of teacher preferences is to recover the forward-looking value of working at each school, V_j, and the other parameters of the search model $(\delta, \lambda_0, \lambda_1, \mathbf{f})$. This is performed for each category of teachers separately, but for clarity in the remainder of this section I suppress the dependence on c in the notation. I start by setting out the identification arguments, before moving on to practical considerations related to estimation.

Identification

The identification of school values in this model rests on relatively simple 'revealed preferences' arguments: if a teacher moves from one school to another, it must be the case that she prefers the new school to the old one; and similarly if a teacher moves from one school to the outside option, it must be the case that she prefers the outside option to the old school. Thus, high-value schools are likely to attract many teachers from other high-value schools, and are unlikely to lose many teachers to low-value schools.

One difficulty, however, is that the model leaves the offer distribution **f** unrestricted, and therefore it is not immediately obvious how school values can be identified separately from offer probabilities. Put simply, if a school is observed to attract many teachers, it could be that it has a high value, or it could be that it sends many offers (eg because it has many vacancies or high recruitment effort). The identifying assumption here is that the offer distribution is the same for teachers working in the inside and outside sectors. Thus, by comparing relative flows from a school to/from the outside option and to/from other schools, it is possible to separately identify school values from offer probabilities.

To understand these identification arguments more formally, let h_{ij} denote the probability of moving from school i to school j, h_{j0} the probability of moving from school j to the outside option, and h_{0j} the probability of moving from the outside option to school j. I assume that these hazard rates are observed in the data.

My structural model implies that:

$$\begin{aligned} h_{ij} &= (1-\delta)\lambda_1 f_j \frac{e^{V_j}}{e^{V_j}+e^{V_i}} \\ h_{0j} &= \lambda_0 f_j \frac{e^{V_j}}{e^{V_0}+e^{V_j}} \\ h_{j0} &= \delta + (1-\delta)(1-\lambda_1)\frac{e^{V_0}}{e^{V_0}+e^{V_j}} \end{aligned} \tag{24}$$

These quantities are invariant to the addition of a constant to the values of all schools and the outside option, so I normalise the value of the outside option to 0 for each category. Then for any two schools i and j we have:

$$\begin{aligned} \frac{h_{0j}}{h_{0i}} &= \frac{f_j}{f_i}\frac{e^{V_j}(1+e^{V_i})}{e^{V_i}(1+e^{V_j})} \\ \frac{h_{ij}}{h_{ji}} &= \frac{f_j}{f_i}\frac{e^{V_j}}{e^{V_i}} \end{aligned} \tag{25}$$

And the ratio between these two quantities is:

$$\frac{h_{0j}}{h_{0i}}\frac{h_{ji}}{h_{ij}} = \frac{1+e^{V_i}}{1+e^{V_j}} \tag{26}$$

This restricts the relationship between the values of any two schools. The level of these values can then be pinned down using the third hazard rate. We have:

$$h_{j0} - h_{i0} = (1-\delta)(1-\lambda_1)\left(\frac{1}{1+e^{V_j}} - \frac{1}{1+e^{V_i}}\right) \tag{27}$$

Denote $\tilde{V}_j = 1/(1+e^{V_j})$. The two expressions above become:

$$
\begin{aligned}
\frac{h_{0j}}{h_{0i}}\frac{h_{ji}}{h_{ij}} &= \frac{\tilde{V}_j}{\tilde{V}_i} \\
h_{j0} - h_{i0} &= (1-\delta)(1-\lambda_1)(\tilde{V}_j - \tilde{V}_i)
\end{aligned}
\tag{28}
$$

For another school k, we have:

$$
\begin{aligned}
\frac{h_{0j}}{h_{0i}}\frac{h_{ji}}{h_{ij}} - \frac{h_{0k}}{h_{0i}}\frac{h_{ki}}{h_{ik}} &= \frac{\tilde{V}_j - \tilde{V}_k}{\tilde{V}_i} \\
\frac{h_{j0} - h_{i0}}{h_{j0} - h_{k0}} &= \frac{\tilde{V}_j - \tilde{V}_i}{\tilde{V}_j - \tilde{V}_k}
\end{aligned}
\tag{29}
$$

Combining these two expressions gives the following identifying correspondence:

$$
\tilde{V}_j = \frac{h_{j0} - h_{i0}}{h_{j0} - h_{k0}} - \left(\frac{h_{0j}}{h_{0i}}\frac{h_{ji}}{h_{ij}} - \frac{h_{0k}}{h_{0i}}\frac{h_{ki}}{h_{ik}}\right)^{-1} \tag{30}
$$

This expression shows that the value of a school is formally identified from relevant hazard rates at that school and at two other schools in the market.

It may seem from this expression that the identification of school values is entirely contingent on school-to-school transitions. This could make estimation difficult since, in any given year, only 5% of teachers undergo such a transition. But the second expression in (29) shows that differences in the hazard rates of moving to the outside option (which is a very common type of transition) are directly informative about the relative values of the schools involved. Moreover the first expression in (25) shows that differences in the hazard of moving from the outside option (which is also a very common transition) combined with some knowledge of the relative value of schools is informative about the relative offer rates of the schools involved. Intuitively, if we observe that a school loses more teachers to the outside option than other schools, it must be the case that this school has a lower value. If we also see that this same school also attracts more teachers from the outside option, it must also be the case that this school 'sends more offers' to teachers: this second observation cannot be explained by a higher value since this has been ruled out by the first observation. Combining these observations across schools allows us to separately identify school values from school offer rates. Thus in practice the estimation of school values is not contingent on there being a large number of school-to-school transitions in the sample.

Estimation

In practice, I estimate all the parameters by maximum likelihood. I assign a categorical variable M_k to each teacher that captures the transition between her states of employment between two years. There are five possible cases: the teacher can move from school j to school i, in which case $M_k = ji$; she can move from school j to the outside option, in which case $M_k = j0$; she can move from the outside option to school j, in which case $M_k = 0j$; she can stay put in school j, in which case $M_k = jj$, or she can stay in the outside section, in which case $M_k = 00$.

The log-likelihood of the sample is then:

$$
\begin{aligned}
LL = \sum_k log\Big[&1[M_k = ji] \times (1-\delta)\lambda_1 f_j \frac{e^{V_i}}{e^{V_j} + e^{V_i}} \\
&+ 1[M_k = j0] \times \left(\delta + (1-\delta)(1-\lambda_1)\frac{e^{V_0}}{e^{V_0} + e^{V_j}}\right) \\
&+ 1[M_k = 0j] \times \lambda_0 f_j \frac{e^{V_j}}{e^{V_0} + e^{V_j}} \\
&+ 1[M_k = jj] \times \left((1-\delta)\lambda_1\Bigg(\sum_{s\in\mathcal{S}\backslash j} f_s \frac{e^{V_j}}{e^{V_s} + e^{V_j}} + f_j\Bigg) + (1-\delta)(1-\lambda_1)\frac{e^{V_j}}{e^{V_0} + e^{V_j}}\right) \\
&+ 1[M_k = 00] \times \left(1 - \lambda_0 + \lambda_0 \sum_{s\in\mathcal{S}} f_s \frac{e^{V_0}}{e^{V_0} + e^{V_s}}\right)\Big]
\end{aligned} \tag{31}
$$

with $\sum_{s\in\mathcal{S}} f_s = 1$ and $1 \geq f_s \geq 0 \quad \forall s$

It is possible to leave the offer distribution completely unrestricted, which involves estimating the probability of getting an offer from each school f_j as part of the maximization of the log-likelihood function (31). However, aside from implying a large parameter space, this approach might be problematic if one wishes to simulate policy counterfactuals that change the size or composition of schools, and offer probabilities happen to be endogenous functions of these school characteristics. Suppose for example that schools with large shares of disadvantaged pupils have higher offer probabilities, for example because they experience higher turnover rates and therefore tend to post more vacancies. If a counterfactual scenario reduces the share of FSM pupils at such schools, using fixed offer probabilities might overestimate the labour supply to these schools.

In principle it would be possible to endogenize the offer distribution in a 'micro-founded' way by making it a function of the number of vacancies posted by each school in each year, which is determined in equilibrium as the left-hand side of the flow-balance equation (9) (more specifically, this is the number of teachers leaving a particular school each year in the steady state). However, this approach would make the computation of the log-likelihood function and its gradient very cumbersome. Also, it involves interpreting differences in offer rates between schools purely as differences in rates of vacancy creations. In reality offer rates might capture other aspects of the job search process, such as search efforts or schools' outreach policies. For these reasons, I use a simpler, 'reduced form' representation of the offer distribution as a function of the size and composition of each school:

$$
f_j = \frac{exp(\lambda_2 q_j + \lambda_3 z_j)}{\sum_{s\in\mathcal{S}} exp(\lambda_2 q_s + \lambda_3 z_s)} \tag{32}
$$

The log-likelihood is invariant to the addition of a constant to all values, so I normalise the value of the outside option to 0 for each category. I assume that the number of teachers in the outside option equals half of the number of teachers in the inside sector. I can then estimate the forward-looking value of each school and the other parameters of the search model, which are shown in Table 13.

Table 13: **Estimates of offer rates parameters**

		Category 1	Category 2
job destruction rate	δ	0.010	0.015
		(0.014)	(0.011)
offer rate (outside)	λ_0	0.138***	0.277***
		(0.004)	(0.005)
offer rate (employed)	λ_1	0.084***	0.135***
		(0.004)	(0.004)
offer distribution (size)	λ_2	0.002***	0.004***
		(0.000)	(0.000)
offer distribution (composition)	λ_3	3.228***	1.454***
		(0.228)	(0.123)
Num.Obs		26234.000	44366.000

The estimates for λ_2 and λ_3 are both positive, implying that teachers are more likely to draw offers from larger, more disadvantaged schools. This is captured in the structural expression for the number of vacancies posted (the left-hand side in equation (9)), and is fairly intuitive: if teachers dislike working for disadvantaged schools, then these schools will experience higher turnover rates and will post more vacancies.

5.5 Utility function parameters

Having estimated the forward-looking values of working at each school, V_j, I use the definition of the value function (6) to recover the mean flow utility of working at each school, ν_j. The normalization I imposed on the forward-looking value of the outside option means that flow utilities are also only recovered up to an additive constant. I assume that the discount rate β is 0.95.

Having recovered the mean flow utility of working at each school, I proceed to recover the preference parameters α^w, α^f. The model implies:

$$\hat{\nu}_j = \alpha^0 + \alpha^w \hat{p}_j + \alpha^f z_j^f + \zeta_j \tag{33}$$

The difficulty is that the observable characteristics of schools - the wage premium they post and the composition of their pupil body - are likely to be correlated with the utility value of their unobservable characteristics: schools that are perceived as desirable for a category of teachers will post a lower wage premium to these teachers; similarly schools that are desirable for a category of teachers may also be desirable for certain categories of pupils.

One possible approach to this issue is to use instruments for the endogenous characteristics of schools (their FSM shares and the premia they post in the labour market). As I am also modelling the 'demand side' of the education market (ie parents and their application decisions), I seek to leverage information on the location and characteristics of households relative to schools. The validity assumption is that the residential location of households relative to schools is independent of the schools' unobserved quality for teachers ζ_j. The idea is that schools located close to a large number of disadvantaged households will tend to enrol larger numbers of disadvantaged pupils and to adjust their wages accordingly as part of their optimization problem in a way that is not driven by their unobserved quality to teachers. As I have an equilibrium model of school choice and wage formation, I can use this information to derive approximations to optimal instruments. I essentially simulate the equilibrium school composition and wages when setting $\zeta_j = 0$, as in section 5.1.

Table 14 below shows the OLS and IV estimates of the parameters in teachers' utility function, for each category of teacher. The OLS estimates suggest that both categories of teachers dislike working in disadvantaged schools, but teachers in category 1 have a slightly stronger aversion for such schools: category 1 teachers request an additional £273 of annual wage for each percentage point increase in the FSM share of a school; whereas for category 2 teachers the corresponding increase is £214 for each percentage point. This is also consistent with the observation that offer probabilities increase faster with the FSM share for category 1 teachers (see the $\lambda 3$ coefficient in table 13): if category 1 teachers dislike disadvantaged schools more, such schools will experience a higher turnover rate and post moe vacancies for such teachers.

Unfortunately, the instruments derived from local geographies are weak, and there are practical issues involved in merging the instruments generated from the 'demand side' with school utilities derived in the 'supply side', resulting in a loss of observations.[12] As a result the IV estimates are very imprecise and not particularly meaningful.

I use a control function approach to mitigate this issue. As part of the analysis of pupils' preferences set out in section 5.1, I estimate the unobserved quality of schools from pupils' point of view (the ξ parameters in their utility function (1)). I assume that the unobservable quality of schools from teachers' point of view can be decomposed as $\zeta_j = \tau^{\zeta}\xi_j + \tilde{\varepsilon_j^{\nu}}$ with $E[\tilde{\varepsilon_j^{\nu}}|p_j, z_j^f] = 0$. This is a strong assumption in the sense that the unobserved quality of schools for teachers and pupils are likely to be imperfectly correlated, and I cannot exclude the possibility that the remaining error term $\tilde{\varepsilon_j^{\nu}}$ is correlated with endogenous characteristics. Nevertheless, this is a plausible way of mitigating the endoegeneity issue. The resulting estimates, also reported in table 14, are similar to the OLS estimates.

[12]These two workstrands are structured as two separate projects in the ONS Secure Research Service, and school identifiers are not always consistent

Table 14: Estimates of parameters in teachers' utility function

	Cat.1 (OLS)	Cat.1 (IV)	Cat.1 (control)	Cat.2 (OLS)	Cat.2 (IV)	Cat.2 (control)
Intercept	-0.607***	-0.052	-0.608***	-0.526***	-0.620***	-0.525***
	(0.042)	(1.248)	(0.042)	(0.025)	(0.190)	(0.025)
School FSM	-0.573***	-0.450**	-0.573***	-0.729***	-0.761	-0.730***
	(0.089)	(0.148)	(0.089)	(0.101)	(0.994)	(0.101)
Wage	0.021***	-0.104	0.021***	0.034***	-0.091	0.034***
	(0.008)	(0.276)	(0.008)	(0.011)	(0.465)	(0.012)
Unobs. quality			-0.001			0.011
			(0.015)			(0.017)
Num.Obs.	367	314	367	371	314	371
R2	0.114	0.05	0.114	0.174	0.066	0.175

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: preference model estimated for classroom teachers employed in West Midlands schools 2017-2019.

5.6 Teacher effectiveness

The next and final step is to estimate the parameters in the education production function (13). Estimating these parameters solely from variations in the composition of school workforces (in the cross-section or over time) is challenging because there are only 351 schools in my sample and the CES functional form involves non-linear parameters. To overcome this issue, I derive a set of additional restrictions from the schools' optimality conditions (15). The basic intuition for this approach is relatively straightforward: my model assumes that schools set wages optimally having observed the preferences and effectiveness of teachers; it follows that, having estimating teachers' wages and preferences, I should be able to recover some information on teachers' effectiveness from the schools' optimality conditions.

This approach is directly inspired from the IO practice of recovering the marginal costs of differentiated products from the firms' profit maximizing conditions (Berry, Levinsohn, and Pakes 1995). However, the application of this approach to my model is not straightforward because schools face a constrained optimization problem: they set the wage applicable to each category of teachers to attract the optimal mix of teachers for their pupil body, subject to not exceeding their budget. As I show below, this structure implies that there is no direct correspondence from the optimality condition for one teacher category to the effectiveness parameter for that category. Instead, the optimality conditions impose some restrictions on the *relative* effectiveness of different teacher categories in each school. Nevertheless, these restrictions can be used jointly with the education production function to support the estimation.

With two categories of teachers, the education production function (13) can be re-written as:

$$\exp(\theta_j) = \gamma_j^0 (\gamma_j^1)^{\frac{1}{\sigma}} q_j^{-1} \left(l_{j1}^\sigma + \frac{\gamma_j^2}{\gamma_j^1} l_{j2}^\sigma \right)^{\frac{1}{\sigma}} \epsilon_j^\theta \tag{34}$$

The value added of schools is not directly observed, but I estimated it as part of my analysis of pupils' choices in section 5.2. I treat it as known here.

The first order conditions of the school's optimization problem with respect to p_{j1} and p_{j2}

are:

$$\begin{aligned}\sigma l_{j1}^{\sigma-1} l'_{j1} - \lambda_j [l'_{j1} \cdot (w^r + p_{j1}) + l_{j1}] = 0 \\ \frac{\gamma_j^2}{\gamma_j^1} \sigma l_{j2}^{\sigma-1} l'_{j2} - \lambda_j [l'_{j2} \cdot (w^r + p_{j2}) + l_{j2}] = 0\end{aligned} \tag{35}$$

Taking the ratio and re-arranging gives:

$$\log\left(\frac{l'_{j1} \cdot (w^r + p_{j1}) + l_{j1}}{l'_{j2} \cdot (w^r + p_{j2}) + l_{j2}} \frac{l'_{j2}}{l'_{j1}}\right) = \log\left(\frac{\gamma_j^1}{\gamma_j^2}\right) + (\sigma - 1)\log\left(\frac{l_{j1}}{l_{j2}}\right) \tag{36}$$

I have estimated all the terms on the left-hand side and the last term on the right-hand side, but in and of itself this restriction does not directly identify the parameters of interest. For each school, this is an equation in two unknowns: the relative effectiveness of the two categories of teachers γ_j^1/γ_j^2, and the parameter σ. To make progress I restrict the heterogeneity in the effectiveness parameters across schools as follows:

$$\gamma_j^1 = e^{\gamma_0^1} z_j^{\gamma_f^1} \quad \gamma_j^2 = e^{\gamma_0^2} z_j^{\gamma_f^2} \quad \gamma_j^0 = e^{\gamma_0^0} z_j^{\gamma_f^0} \tag{37}$$

That is, the effectiveness of a category of teacher in a school is a function of a constant and the share of disadvantaged pupils enrolled in that school. This allows for 'matching effects' between teachers and school types, albeit in a more restricted way.

Equations (34) and (36) then become:

$$\begin{aligned}\theta_j - \log(q_j^{-1}) = \tilde{\gamma}_0^0 + \tilde{\gamma}_f^0 \log(z_j) + \frac{1}{\sigma}\log\left(l_{j1}^{\sigma} + e^{-\tilde{\gamma}_0^1} z_j^{-\tilde{\gamma}_f^1} l_{j2}^{\sigma}\right) + \tilde{\varepsilon}_j^{\theta} \\ \log\left(\frac{l'_{j1} \cdot (w^r + p_{j1}) + l_{j1}}{l'_{j2} \cdot (w^r + p_{j2}) + l_{j2}} \frac{l'_{j2}}{l'_{j1}}\right) = \tilde{\gamma}_0^1 + \tilde{\gamma}_f^1 \log(z_j) + (\sigma - 1)\log\left(\frac{l_{j1}}{l_{j2}}\right) + \varepsilon_j^{FOC}\end{aligned} \tag{38}$$

where the following transformations of the parameters must be estimated:

$$\tilde{\gamma}_0^0 = \gamma_0^0 + \frac{\gamma_0^1}{\sigma}, \quad \tilde{\gamma}_f^0 = \gamma_f^0 + \frac{\gamma_f^1}{\sigma}, \quad \tilde{\gamma}_1^0 = \gamma_0^1 - \gamma_0^2, \quad \tilde{\gamma}_1^f = \gamma_f^1 - \gamma_f^2 \tag{39}$$

The error term in the second equation can be interpreted as the effect of measurement error and/or as the effect of unobserved heterogeneity in the effectiveness parameters. For example, if instead of the restriction assumed in (37) we have $\gamma_j^1 = e^{\gamma_0^1} z_j^{\gamma_f^1} e^{\epsilon_j^1}$ and $\gamma_j^2 = e^{\gamma_0^2} z_j^{\gamma_f^2} e^{\epsilon_j^2}$, then we have $\varepsilon_j^{FOC} = \epsilon_j^1 - \epsilon_j^2$. This raises the possibility that this error term is correlated with some of the terms in this second equation, notably l_{j1}/l_{j2}: if a school observes that a category of teacher is particularly effective for its operating conditions, then it is likely to seek to recruit more teachers of that category.

The first equation also suffers from a potential endogeneity problem operating through teachers' preferences and labour supply decisions. If the unobserved determinants of school effectiveness captured in ε_j^{θ} are correlated with the unobserved determinants of teachers preferences captured in ζ_j in equation (4), then they might be correlated with teacher numbers l_{j1} and l_{j2}.

As for the estimation of preferences discussed above, there are two potential solutions to this problem. The first approach is to use information on the relative location of households and schools to derive instruments for the number of teachers employed by each school in each category. The idea is that schools located close to a large number of disadvantaged households are likely to enroll a larger proportion of disadvantaged pupils and are likely to find it optimal to adjust their optimal pay and hiring decisions accordingly. The second approach is a control function model using estimates of the unobserved quality of schools ζ_j as a proxy for the omitted variables in equations (38).

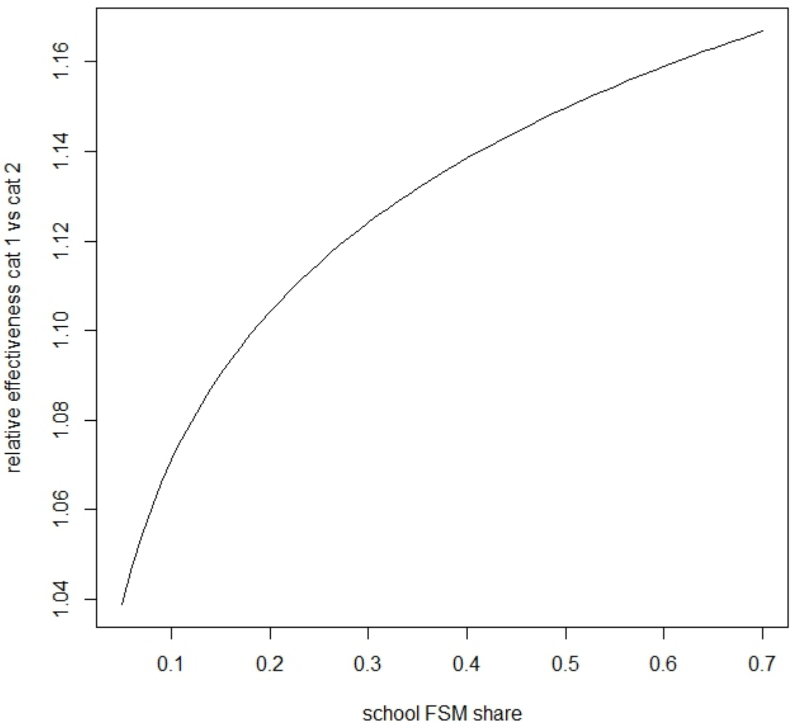
Table 15 below reports the results estimated by GMM. As for the estimation of preferences, the IV estimates are very imprecise, and the control function estimates closely resemble the OLS estimates. The estimates of γ_0^1 and γ_f^1 are both positive, indicating that category 1 teachers tend to be more effective than category 2 teachers, and that the gap in effectiveness increases with the share of FSM pupils employed by schools. That is, all schools will find it more advantageous to hire more category 1 teachers, but more disadvantaged schools even more so. Figure 5 below graphs the relative effectiveness of category 1 vs category 2 teachers (that is $\gamma_j^1/\gamma_j^2 = \tilde{\gamma}_0^1 + \tilde{\gamma}_f^1 \log z_j$) over the range of observed FSM shares.

Table 15: Estimates of parameters in production function

	Baseline	Instruments	Control function
gamma^0_0	1.273***	1.357***	1.201***
	(0.041)	(0.065)	(0.138)
gamma^0_f	-0.102***	-0.090***	-0.100***
	(0.020)	(0.034)	(0.020)
gamma^1_0	0.178***	0.215***	0.170***
	(0.024)	(0.035)	(0.024)
gamma^1_f	0.047***	0.044	0.044***
	(0.014)	(0.046)	(0.014)
sigma	1.210***	1.323***	1.209***
	(0.018)	(0.146)	(0.018)
control_z1			-0.279
			(0.193)
control_z2			0.168
			(0.159)
Num.Obs.	355	316	355

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5: **Relative effectiveness of Cat.1 vs Cat.2 teachers**



One potential concern is that these results might be driven largely or completely by the additional restrictions derived from the schools' optimality conditions. As a sense check, I also consider whether the patterns set out in table 15 also hold when these restrictions are not applied. To this end, I estimate the parameters of a simpler, linear model using the whole sample of schools and the Progress 8 metric as a proxy for value added.[13] As explained in section 2, the funding policy applied in the period considered implies that the budgets of LAs were largely based on historical factors rather than their current operating conditions, and this is likely to induce a degree of variation in teachers-to-pupil ratios in the cross section that is not purely reflective of factors that could be correlated with the error term in the production function. Table 16 shows that the patterns identified in the CES model also hold in this simpler model - ie teachers in category 1 are more effective, and their effectiveness decreases at a lower rate with the share of FSM pupils in schools

[13]The model I estimate is $Progress8_j = \gamma_0 + \gamma_1^0 \frac{l_{j1}}{q_j} + \gamma_1^f \frac{l_{j1}}{q_j} z_j^f + \gamma_2^0 \frac{l_{j2}}{q_j} + \gamma_2^f \frac{l_{j2}}{q_j} z_j^f + \epsilon_j$

Table 16: **OLS estimates of linear production function parameters for secondary schools in England**

	Model 1
Intercept	-0.082
	(0.060)
cat 1 teacher-pupil ratio	8.069***
	(1.455)
cat 1 teacher-pupil ratio x school FSM	-0.056**
	(0.029)
cat 2 teacher-pupil ratio	6.548***
	(1.205)
cat 2 teacher-pupil ratio 2 x school FSM	-0.359***
	(0.031)
Num.Obs.	2959
R2	0.200
R2 Adj.	0.199
se_type	HC2

* p < 0.1, ** p < 0.05, *** p < 0.01

6 Policy counterfactuals

Having developed and estimated a model of teachers-to-schools matching and a model of pupils-to-schools matching, I can simulate counterfactual allocations and outcomes for different policies. For this purpose I assume that the principal policy lever available to policy makers operates through school funding, and more specifically an additional budget allocation made for each percentage of FSM pupils (the 'pupil premium'). I further assume that policy makers are primarily interested in educational inequality, and that their objective can be formulated simply as reducing the mean attainment gap between affluent and disadvantaged households.

To simulate counterfactual prices, allocation, and educational outcomes, I need to define the equilibrium concept applicable to this model. A sorting equilibrium is defined by three objects: the ordered lists submitted by households $\{L_i\}_{i\in\mathcal{P}}$; the wages set by schools for the different categories of teachers $\{\mathbf{w}_j\}_{j\in\mathcal{S}}$; and an allocation of teachers to schools $\{\mathbf{t}_j\}_{j\in\mathcal{S}}$. The tuple $\left\{\{L_i\}_{i\in\mathcal{P}}, \{\mathbf{w}_j\}_{j\in\mathcal{S}}, \{\mathbf{t}_j\}_{j\in\mathcal{S}}\right\}$ is a rational-expectations sorting equilibrium if the following conditions are satisfied:

- the ordered lists $\{L_i\}_{i\in P}$ reflect households' consideration sets and preferences as defined in (1);
- the wages set by schools $\{\mathbf{w}_j\}_{j\in S}$ solve their optimisation problem defined in (14);
- the allocation of teachers to school $\{\mathbf{t}_j\}_{j\in S}$ reflects the preferences given in (4); and
- expectations are rational, that is $\hat{q}_j = q_j$, $\hat{z}_j^f = z_j^f$, and $\hat{z}_j^p = z_j^p$.

Simulating this equilibrium is computationally intensive, and it is not currently possible to perform this task in the ONS Secure Research Service (SRS), which is the environment used for this analysis. For this reason, I perform this exercise on a fictitious economy simulated outside the SRS to reflect the key characteristics observed for the secondary education system in the West Midlands. These characteristics are: the overall share of disadvantaged households, the distribution of disadvantaged households across and within local authorities, the average distance to schools listed in applications, the covariance between FSM status and primary school test scores, the attainment gap between affluent and disadvantaged pupils, and the number and composition of the teaching workforce (both employed and unemployed). I assume that schools simply rank applying pupils based on distance (which is the most common criterion used to break ties after coarser criteria have been applied).

The computation procedure essentially looks for a fixed point in the endogenous school characteristics (FSM share and AC5EM score) that satisfies the equilibrium conditions set out above. It involves the following steps:

- Step 0: specify arbitrary starting values of school endogenous characteristics;
- Step 1: compute the utility of being assigned to each school for each household according to (1) and rank the 5 closest schools to each household accordingly;
- Step 2: run the Gale-Shapley deferred acceptance algorithm to obtain the allocation of pupils to schools, and compute the resulting size and FSM share of each school;
- Step 3: compute the resulting budget allocation of each school;
- Step 4: find equilibrium wages and allocations in the teacher labour market by simultaneously solving the system of optimality conditions (15);
- Step 5: find the resulting value added of each school based on the education production function (13);
- Step 6: find the resulting test score of each individual pupil based on the test score production function (2), and compute the corresponding AC5EM score of each school;
- Step 7: compare the resulting school characteristics with the starting values, and re-iterate from Step 1 if the difference exceeds a tolerance.

Table 17 below reports the mean attainment gap between affluent and disadvantaged pupils, as well as the mean teacher-to-pupil ratio at schools in the most disadvantaged quintile. This shows that even large increases in the pupil premium only have a fairly modest impact on educational inequality: doubling the pupil premium from £1k to £2k only reduces the attainment gap from 0.612 standard deviation to 0.607 standard deviation. Increasing the pupil premium to £5k reduces the attainment gap to 0.595 standard deviation. Even at this higher levels, the pupil premium only represents a share of the total per pupil funding granted to schools.[14] Moreover, my analysis of the determinants of test scores set out in section 5.2 shows that

[14] By way of reference, the total wage bill divided by the number of newly enrolled pupils in 2019 was £12k per year

differences in school value added only account for a share of differences in outcomes. Lastly, improvements in school staffing resulting from better funding of disadvantaged schools benefits both disadvantaged pupils and the affluent pupils who are educated alongside them, such that the attainment gap only reduces by a modest amount.

Table 17: **Counterfacual policy simulations**

Premium (k£ per pupil)	**Attainment gap (% of SD)**	**Pupil to teacher ratio (cat 1)**	**Pupil to teacher ratio (cat 2)**
1	61.2%	0.121	0.221
2	60.7%	0.124	0.226
3	60.3%	0.128	0.229
4	59.9%	0.132	0.233
5	59.5%	0.134	0.237

7 Conclusion and discussion

This paper develops and estimate a structural, 'two-sided' model of the school system in England. The empirical analysis shows that different types of households make different trade offs when applying to schools: disadvantaged households put more weight on distance (ie face higher travel costs), and less weight on school performance and school composition; and conversely, households with higher primary test scores put less weight on distance, and more weight on school performance and on school composition. These patterns of heterogeneity imply that, in and of itself, school choice may not necessarily reduce inequality in attainment. Pupils from more affluent families are more likely to use the scheme to seek admission to better performing schools, and this is likely to exacerbate inequality in attainment in secondary education.

This analysis also indicates that pupils' unobserved ability correlates with their unobserved preferences for school quality - in other words pupils who put more weight on school performance are also likely to do better in secondary school exams than suggested by their observable characteristics. This implies that measures of school value-added that control solely for observable pupil characteristics are imperfect proxies of true value-added. However, such simple measures of value added do not seem to systematically understate differences in the true value added between schools.

My estimation of preferences relies on the assumption that households' applications truthfully reveal their preferences, ie households rank all the schools they consider, and rank these schools in the true order of preferences. This assumption may be violated if households face costs when ranking schools they have considered (which might apply to the 18% of households in my sample who rank the maximum number of schools allowed in their LA), or if households face no uncertainty in their probability of admission in certain schools (which could apply for example to some households located very far away from oversubscribed schools that they like). A breach of this assumption is likely to bias the coefficient on school performance downward, as some households do not rank very good schools that they do in fact find desirable.

In principle, it might be possible to address this issue by exploiting the 'stability' property of the deferred acceptance mechanism (Fack, Grenet, and He 2019). This property implies that every pupil is matched with her favorite school among those she qualifies for ex post. If the researcher observes the ranking of schools over pupils and the cutoff applicable at each school, then the researcher can essentially delineate the 'choice set' of each pupil (the set of schools for which that pupil qualified) and use standard discrete choice models to recover preferences. Unfortunately this approach is not feasible in England as although schools publish their prioritization criteria, the NPD does not contain information on the attributes of pupils that would inform their ranking by schools.

The model of the job market used in this paper combines some aspects derived from the labour literature on job search, with other features derived from the IO literature on differentiated product markets. In this model, the market power of employers (schools) comes from the fact that workers cannot instantly receive an offer from all schools (the search frictions), but also from the fact that workers have preferences over non-wage attributes of jobs, and therefore might be willing to work for desirable schools that pay less than competitors. This concept is often referred to as 'compensating differential' in the labour literature (eg Antos and Rosen 1975). The model also allows teachers' effectiveness to vary across teacher categories and school types, consistent with a Roy model of labour allocation.

The analysis confirms that most teachers tend to dislike working for schools with large shares of disadvantaged pupils, but there is significant heterogeneity among teachers in both preferences and effectiveness. One category of teachers appears to have a slightly stronger distaste for working in disadvantaged schools, but is also relatively more effective than other teachers in such schools. These teachers tend to be paid more than other teachers, especially by more disadvantaged schools, and are more likely to work in such schools. These results suggest that the observed tendency for more disadvantaged schools to hire greater shares of inexperienced teachers (Table 2) cannot be explained by the fact that such teachers have a preference for working in such schools or are more effective in such environments. Instead, it simply reflects the dynamics of job search - in other words, inexperienced teachers are more likely to work for disadvantaged schools simply because such schools tend to 'send more offers' (likely because they experience stronger turnover rates) and inexperienced teachers have not received offers from 'better schools' yet.

The process of wage formation applicable to teachers in England operates under complex regulatory constraints encapsulated in the pay scales published by the DfE. To make the model tractable, I assume that the wage is simply the sum of a regulated component that varies with experience (in line with the pay scales), and a flexible 'premium' that is applicable to all teachers within a category irrespective of experience. If more experienced teachers were not in fact more effective, and if schools effectively used their powers to flex wages depending on teacher effectiveness, they might face an incentive to 'flatten' the pay curve with respect to experience.[15] In practice they do not appear to do this to any significant extent, in the sense that the gradient of pay with respect to experience does not seem to have changed after the introduction of the wage reforms in 2013. Also, recent research indicates that teachers'

[15] In the model this would take the form of a pay premium that would vary negatively with wages

effectiveness does increase with experience (Kini and Podolsky 2016), so it seems unlikely that this incentive would be very strong.

I also assume that all schools (whether academies or maintained) have the same degree of flexibility with respect to wage-setting. This is a strong assumption, but I show that a simpler empirical model that does not use such restrictions yields similar results with respect to the relative effectiveness of teachers (see section **??**).

In this model, the only variable flexed by schools is the wage (or more specifically, the wage premium paid above regulated wage). The model 'endogenizes' the offer distribution (the probability that a teacher draws an offer from each school) by making it a function of observable school characteristics, but it is not a decision variable. Further research could set out a more complex model of vacancy formation and hiring decisions.

References

[Abd+20] Atila Abdulkadiroglu et al. "Do Parents Value School Effectiveness?" eng. In: *NBER Working Paper Series* (2020).

[AGN19] Claudia Allende, Francisco Gallego, and Christopher Neilson. "Approximating the equilibrium effetcs of Informed School Choice". In: *Working paper* (2019).

[All19] Claudia Allende. "Competition under social interactions and the design of education policies". In: *Job Market Paper* (2019).

[ALM24] Francesco Agostinelli, Margaux Luflade, and Paolo Martellini. *On the Spatial Determinants of Educational Access.* Working Paper 32246. National Bureau of Economic Research, 2024. DOI: 10.3386/w32246. URL: http://www.nber.org/papers/w32246.

[AR75] Joseph R Antos and Sherwin Rosen. "Discrimination in the market for public school teachers". eng. In: *Journal of econometrics.* Journal of Econometrics 3.2 (1975), pp. 123–150. ISSN: 0304-4076.

[Bat+22] Michael D Bates et al. "Teacher Labor Market Equilibrium and Student Achievement". eng. In: *NBER Working Paper Series* (2022). ISSN: 0898-2937.

[BJL16] Stéphane Bonhomme, Grégory Jolivet, and Edwin Leuven. "School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities". In: *Economic Journal* 126.594 (2016), pp. 1342–1371. ISSN: 0013-0133.

[BLM19] Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. "Discretizing Unobserved Heterogeneity". In: *Unpublished Manuscript* (2019).

[BLM22] Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. "Discretizing Unobserved Heterogeneity". eng. In: *Econometrica* 90.2 (2022), pp. 625–643. ISSN: 0012-9682.

[BLP95] Steven Berry, James Levinsohn, and Ariel Pakes. "Automobile Prices in Market Equilibrium". eng. In: *Econometrica* 63.4 (1995), pp. 841–890. ISSN: 0012-9682.

[BM98] Kenneth Burdett and Dale T. Mortensen. "Wage Differentials, Employer Size, and Unemployment". eng. In: *International economic review (Philadelphia)* 39.2 (1998), pp. 257–273. ISSN: 0020-6598.

[Boy+13] Donald Boyd et al. "Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers". eng. In: *Journal of labor economics* 31.1 (2013), pp. 83–117. ISSN: 0734-306X.

[Bur+15] Simon M Burgess et al. "What parents want: School preferences and school choice". eng. In: *The Economic journal (London)* 125.587 (2015), pp. 1262–1289. ISSN: 0013-0133.

[Cha87] Gary Chamberlain. "Asymptotic efficiency in estimation with conditional moment restrictions". eng. In: *Journal of econometrics.* Journal of Econometrics 34.3 (1987), pp. 305–334. ISSN: 0304-4076.

[CTT22] Julien Combe, Olivier Tercieux, and Camille Terrier. "The Design of Teacher Assignment: Theory and Evidence". eng. In: *The Review of economic studies* 89.6 (2022), pp. 3154–3222. ISSN: 0034-6527.

[Dep17] DepartmentForEducation. "Evaluation of Teachers' Pay Reform". In: 133 (2017).

[DF18] L. E Dubins and D. A Freedman. "Machiavelli and the Gale-Shapley Algorithm". eng. In: *The American mathematical monthly* 88.7 (2018), pp. 485–494. ISSN: 1930-0972.

[DI10] Steven N Durlauf and Yannis M Ioannides. “Social Interactions”. eng. In: *Annual review of economics* 2.1 (2010), pp. 451–478. ISSN: 1941-1383.

[DK02] Stacy Berg Dale and Alan B Krueger. “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables”. eng. In: *The Quarterly journal of economics* 117.4 (2002), pp. 1491–1527. ISSN: 0033-5533.

[Ede22] Tim Ederer. “Labor market dynamics and teacher spatial sorting”. In: *Job Market Paper* (2022).

[FGH19] Gabrielle Fack, Julien Grenet, and Yinghua He. “Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions †”. eng. In: *American Economic Review* 109.4 (2019), pp. 1486–1529. ISSN: 0002-8282.

[Hec79] James J. Heckman. “Sample Selection Bias as a Specification Error”. eng. In: *Econometrica* 47.1 (1979), pp. 153–161. ISSN: 0012-9682.

[JPV20] Gregory Jolivet and Fabien Postel-Vinay. “A structural analysis of mental health and labor market trajecories”. In: *Working Paper* (2020).

[KP16] T. Kini and A. Podolsky. “Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research”. In: *Palo Alto: Learning Policy Institute* (2016).

[MM23] Christian Moser and Iacopo Morchio. “The gender pay gap: micro sources and macro consequences”. In: *Working paper* (2023).

[Pak21] Ariel Pakes. “A Helicopter Tour of Some Underlying Issues in Empirical Industrial Organization”. eng. In: *Annual review of economics* 13.1 (2021), pp. 397–421. ISSN: 1941-1383.

[RV14] Mathias Reynaert and Frank Verboven. “Improving the performance of random coefficients demand models: The role of optimal instruments”. eng. In: *Journal of econometrics* 179.1 (2014), pp. 83–98. ISSN: 0304-4076.

[SFB21] John R Stromme, Chao Fu, and Barbara Biasi. “Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage”. eng. In: *NBER Working Paper Series.* NBER Working Paper Series (2021). ISSN: 0898-2937.

[Sor18] Isaac Sorkin. “Ranking Firms Using Revealed Preference”. eng. In: *The Quarterly journal of economics* 133.3 (2018), pp. 1331–1393. ISSN: 0033-5533.

[Tin21] Michela Tincani. “Teacher Labor Markets, School Vouchers and Student Cognitive Achievement: Evidence from Chile”. eng. In: *Quantitative Economics , 12 pp. 173-216. (2021)* (2021).

[Tra09] Kenneth E. Train. *Discrete Choice Methods with Simulation / Kenneth E. Train.* eng. 2nd ed. 2009. ISBN: 9780511805271.

[TW03] Petra E Todd and Kenneth I Wolpin. “On the specification and estimation of the production function for cognitive achievement”. eng. In: *The Economic journal (London)* 113.485 (2003), F3–F33. ISSN: 0013-0133.

[Wel18] Matthew Weldon. “Secondary school choice and selection - Insights from new national preferences data”. In: *Research report prepared for the Department of Education* (2018). ISSN: 0013-0133.