LUND UNIVERSITY
School of Economics and Management

Master's Programme in Finance

Investigating the effect of good schools on surrounding housing prices in London

by

Leslie Shengnan Liu and Nicholas Sampson

NEKN02
Master's Thesis (15 credits ECTS)
June 2018
Supervisor: Kaveh Majlesi

Abstract:

In this paper we have looked at the effect good schools have on surrounding house prices, focusing on London. We consider both the effect of "% of students that meet expected standard in math, reading, and writing", and the new "progression score" that the UK government has introduced. These two scores are used as the basis of how a school is defined as 'good'. We used two different propensity score matching methods to estimate the effect that good schools have on surrounding house prices. We got different results depending on which matching method we used, and which score to define good schools. "% of students that meet expected standard in math, reading and writing" showed a positive effect for both the Nearest Neighbor method (8.42%) and the Kernel method (3.74%) whereas "progression scores" showed a negative effect for the nearest neighbor method (-9.95%) and positive for the Kernel method (0.59%).

Acknowledgements

We would like to thank everyone who supported us at the university and a special thanks to our supervisor Dr. Kaveh Majlesi for his constructive and helpful comments.

Table of Contents

List of Tables

List of Figures

1. Introduction

In this paper we are trying to find out how much house prices are affected by good schools within the neighborhood. The paper will investigate the difference between average housing prices in a neighborhood, in the case whereby it has a good school as compared to having a bad school.

We decided to focus on London as we believe it will be an interesting city to look at. London is home to many top tier schools and well known for producing highly educated talents. It is also a city with an extremely varied demographic composition, and leading in arts, education, entertainment, tourism and finance; making it a favorite option of investors from all over the world. The 2017 Brexit could also potentially influence the housing market in United Kingdom in the next decade, making London a unique city to investigate for our study.

Previous papers on this subject that looked at various cities/countries around the world has found that one standard deviation change in test scores will increase house prices between 2% to 4%. Most papers have built on Black's (1999) paper which looks at house prices within the same neighborhood but on opposite sides of the school attendance boundary. By looking at house prices in the same neighborhood but in different school attendance boundaries, she was able to control for neighborhood characteristics. This research method will work in cities where school attendance boundaries exist and are stable but not in London where school attendance boundaries are primarily based on the distance among the students that are applying that year.

The determinants of school quality have been widely researched as many researches have tried to determine what parents value and consider a good school. The most commonly suggested factor that parents care about are test scores, but the additional value a school provides (the additional learning the students get from the teaching) has also been looked at. Other factors such as ethnicity of students, reputation and leadership/teaching style of the school have also been mentioned.

As London do not have set school attendance boundaries, we opted for a different methodology compared to Black's (1999) paper. We have used a propensity score matching method where we match a good and a bad school located in similar neighborhoods. The neighborhoods are defined based on the 2011 Census Survey which provides demographic information on the UK. Then, we control for neighborhood characteristics by matching neighborhoods with similar

characteristics using propensity score. In the paper, a good school is defined and classified using two methods – “progress scores” and “% of students that meet expected standard in math, reading and writing” over 2 years”.

We believe that this is an important topic to investigate as it will be of good use to economists, educators and policy makers. In the eyes of educators, this will show the connection that they work will have on the surrounding house prices. In the field of economics, this will help further contribute to the understanding of the relationship between school performance and housing prices in the vicinity. To policy makers, this study could help quantify the tangible effects of increasing governmental budget on improving the quality of education of schools. Moreover, it will show how well the latest publicly available school score formats provided by the UK government relates to the school and house price relationship. This research could ultimately be valuable to governmental organizations when creating policies that involve making decisions concerning matters related to housing or schools.

2. Literature Review

2.1. Earlier papers on the connection between school quality and house prices

Much research has been done on the connection between school quality and house prices. The basic format of earlier papers has been to regress house prices against school test scores while controlling for house and neighborhood characteristics.

However, this normally comes with the endogeneity problem, where better schools are normally located in wealthier areas and students from wealthier backgrounds normally do better academically. Failing to correct for observable and unobservable characteristics which might be correlated with both house price and school performance, could result in biased estimates. (Fack & Grenet, 2008)

One of the most quoted paper, Black's (1999) paper looked at how house prices in Massachusetts changed when they are part of the catchment area of a good school. Black looked at house prices within 0.15 miles from the attendance boundary of the catchment areas, in neighborhoods split by this boundary, as you will have children from the same neighborhood going to different schools. Thus, controlling for neighborhood characteristics. After the neighborhood effect has been controlled for, the price differences of similar houses on the opposite side of the boundary would be a result of school quality. Black found a 5% increase in test scores in primary school test resulted in a 2.1% ($3948) increase in house price. This was half compared to a "naïve" OLS regression, where Black only regress house prices against test scores. (Black S. E., 1999)

However, the problem with this model is that house prices close to the boundary might be lower compared to houses closer to the school, due to possible future changes in the location of the boundary. (Yinger & Nguyen-Hoang, 2011)

Stephen Gibbons' (2012) paper which built on Black's (1999) paper, found similar results looking at data from UK using an improved boundary discontinuity regression methodology by matching identical properties across authority boundaries. They found that one standard deviation change in school average value-added or prior achievement, increases the price close to 3%. (Gibbons, Machin, & Silva, 2012)

Other papers focusing on different areas around the world have come to a similar conclusion that one standard deviation change in test scores will result in a 3% to 4% change in house prices. (Machin & McNally, 2008)**,** (Black & Machin, 2010) and (Machin, 2011)**.**

This has been further supported by Nyguyen-Hoang and Yinger's (2011) paper who did a review on various papers looking at this problem by examining methodologies and capitalization reviews. They found that on average one standard deviation change in test scores increased house values by below 4%. (Yinger & Nguyen-Hoang, 2011)

2.2. School Quality

When it comes to school quality researchers have not just looked at test scores, but also considered the "value added" - the additional learning that the school contribute (Black & Machin, 2009). It comes down to what parents find important when choosing a school and what factors they consider. The problem with using test scores is that better test scores can be a result of improved enrolment quality or greater pupils' progress, not necessary better school performance.

Gibbons' (2012) paper considered what parents deem important when deciding on which school to send their children too. The paper looked at the scenario where parents pay for the "expected academic gains for their children" (value added), or "good peers and school composition regardless of how these factors affected their child's school achievements". The results indicated that house prices responded equally to expected academic gains and the initial characteristics of students, meaning that parents care about both (Gibbons, Machin, & Silva, 2012).

Figlio and Lucas' (2004) paper looked at how state-administered school grades (a grade for the school, not individual students) effected house prices in Florida. They found that the effect was most prominent in the years after 2002, when the grading system was first introduced. However, due to the volatility of the school grades, they stated that the informative power of the grades diminished over time. Although, "A" schools that consistently performed well were shown to increase house prices over several years (Figlio & Lucas, 2004). The Office of Standard in Education, Children's Service and Skills (Ofsted) rating could be considered a similar rating in UK. Ofsted regulates services that care for children of all ages. They carry out hundreds of inspections throughout the year, with the goal to achieve excellence in education and skills for

learners of all ages, and in the care of children and young people. They are independent and impartial and report directly to the UK Parliament (Ofsted, 2018).

Schneider and Buckley looked at parental preferences by studying information search patterns and compare these findings to the relevant literature. They found that the ethnicity and economic background of pupils were most important to parents, which contradicted their survey results. This would suggest that parents care more about demographic composition than they would like to admit publicly. Test scores were also accessed in high numbers, but information on teacher quality was not highly visited (Schneider & Buckley, 2002).

This was further supported by Clap (2007) who found that student ethnicity was a deciding factor for parents. His study looked whether property buyers pay for test scores or demographic composition, based data on spanning from 1994 to 2004. The study showed that an increase in the percentage of Hispanic students had a negative effect on house prices, but the effect of test scores was mixed. Nonetheless, the study also showed that the negative effect of percentage of Hispanics was declining and the importance of test scores were increasing over time (Clapp, Nanda, & Ross, 2007).

The Organisation for Economic Co-operation and Development (OECD) found that parents value safety and reputation more than academic achievement, in all the eight countries that they investigated. However, UK was not one of the countries included (OECD, 2015).

However, National Foundation of Education Research (NEFR) conducted similar research in UK, (reputation and safety were not included as a factor). They showed that “School that most suits my child”, “Location”, and “Discipline/behavior that promotes effective learning”, were the top three factors. Having said this, there was a disparity between lower and higher income families. Higher income families placed higher value on “Discipline/behavior that promotes effective learning”, “Examination results” and “Effectiveness of school’s senior leadership team”, where lower income families placed higher value on “Location”, “Well-qualified teachers” and “Reputation of taking parents or careers” views into account. Middle class families were in the middle for all results (Wespieser, 2018). From past research we can see that it is hard to pin down exactly what factors parents value the most when deciding on schools.

The UK government publish statistics on state controlled primary schools every year. The current main measure that the UK government publicizes is called “progression score measure”, (Standar & Testing Agency, 2017).

This is a relatively new ranking system, as it was introduced in 2016, and two years of data is available. We have yet found a paper with UK data that uses the “progression score” as an indicator of school quality.

Based on our research we have decided to look at both “% of students that meet expected standard in math, reading and writing” and the “progression score” as an indicator of school quality, to cover both the traditional test scores and “value added”. Moreover, as ethnicity has been highlighted as a deciding factor, we decided to include that as well, among other traditional demographic factors, discuss in more detail in the next section.

3. Data

3.1. Type of Schools and school quality

We decided to focus on primary schools' test scores between 2016 and 2017. We are focusing on schools in London, not the whole of UK. To be more exact on Greater London (often referred to as London) which is a region in UK consisting of 32 boroughs, covering 1,572 sq. km and a population of 8.9 million people (LondonDatastore, 2018).

Nursery (Key Stage 1; 5-7 years old), Secondary schools (Key Stage 3 and 4 ;11 to 14 years old) and colleges (Key Stage 5; 16 to 18 years old) have been excluded and we only include primary schools (Key Stage 2; 7 to 11 years old).

Among primary schools the most common type of schools are academies, community schools and free schools. We are only looking at community schools as academic and free schools are not fully controlled by Local Authorities (LAs) or boroughs (GOV.uk, 2018). Thus, does not always have to follow the admission policies set by the LAs. For example, religious schools (included in free schools) base part of their decision on the religion of the student. Whereas Academies are their own admission authority, but they still have to follow the School Admission code, that came in 19 December 2014.

In total this would add up to 813 community primary schools for us to analyze.

All the data on basic school information was collected from GOV.UK a governmental website. The site provided information such as school name, postcode, the school unique reference number (URN), type of school, age rage, which LA it was located in etc. This is all publicly available information.

We decided to focus on primary schools, as we believed that if someone were to move to an area because of a good school, they most likely do so for a primary school, as their admission is based on distance from school and it is a first stage of a child's education. To be more specific the primary school admission criteria for community schools are based on "distance from school" and "if the student had/have a sibling in the school". However, there are exceptions for educational need of the child in the following situations, "Education, Health & Care plan which names the school", "Looked after & previously looked after children", "Children with a child Protection Plan" and "Medical/Social reasons" and in some cases Child of staff.

As distance from school is the only thing that parents can control, intuitively one can assume that parents will pay a premium to live close to a “good” primary school.

Therefore, we decided to focus on primary schools.

3.2. Test scores/school quality

Key Stage 2 Score is a national test that students take at the end of primary school. Every year the Department for Education (DfE) provide statistical information on how well the students performed.

The current measure that is used as an indicator of school performance is what’s called “progress score”. Rather than just averaging the score, the “progress score” measures the progress the students have done in reading, writing and math, between Key Stage 1 (usually at age 7) and Key Stage 2 (usually at age 11) compared to students across the country with the same Key Stage 1 scores. A score above zero would mean that students have improved more on average compared with other students, with similar Key Stage 1 score.

A progress score below zero, would mean that the students on average have progressed less then students with similar Key Stage 1 scores. Most schools have progression scores between -5 and +5 (DfE, 2016).

Another measure that we used is the “% of student that meet the expected standard in reading, writing and math”. Students are measured on a scaled score based on their raw score. Raw score meaning the total marks they get on the exam. A scaled score of 100 represent the expected standard. Meaning that if a student scores 100 or above they have meet the expected standard. However, as the questions change every year, this means that the difficulty can change to some extent year-on-year. Therefore, the score needed to reach the expected standard may change slightly over the years, depending on whether the test is deemed easier or more difficult. (Standards and Testing Agency, 2017)

Both these scores are only available for 2016 to 2017, as the DfE used different methods to compare schools before 2016. We tried to find similar data release prior to 2015, to enable us to analyze the further back then 2016, by calling and emailing DfE. However, over the phone we were told that the school performance measure has changed several times since 2011 (the year our demographic information starts from) and it would be hard to compare back further

then 2016. We didn't get any response to our email. Comparing school performance datasets back to 2011, we couldn't see any important measure that tracked back to 2011 after 2016.

The "progression score" and the "% of student that meet the expected standard in reading, writing and math" were collected in the same place as the school information, described above, and we were able to match the two excel sheets through URNs.

We believe that progression scores include an element of both value added and test scores. Test scores is what parents probably value the most, based on our empirical research. However, it also includes an element of "value added", which intuitively one can assume is one of the most important factors. Additionally, this score is what is promoted and easiest to find, when you go on the governments website to compare schools.

Overall performance at end of key stage 2 in 2017 - all pupils

School name	Type of school	% of pupils meeting expected standard	Progress score & description			% of pupils achieving at a higher standard	Average score in reading	Average score in maths
			Reading	Writing	Maths			
St Clement Danes CofE Primary School	Maintained School	82%	Well above average 4.0	Average 1.6	Well above average 5.1	11%	107	108
St Matthew's School, Westminster	Maintained School	64%	Average 0.2	Average 1.7	Average -0.3	4%	104	104
Soho Parish CofE Primary School	Maintained School	55%	Average 0.0	Average -2.1	Below average -2.8	14%	105	102
St Josephs Catholic Primary School	Maintained School	52%	Below average -2.7	Average -0.4	Well below average -3.9	4%	101	99

Figure 3.1: How schools are presented when parents compare schools on GOV.UK

In the figure above, we can see that "progression scores" are presented in a more visual appealing way. This makes us assume that new parents to this site, will put more importance on the progression score. Whereas, "% of student that meet the expected standard in reading, writing and math" is probably intuitively easier to understand, therefore we looked at this measure as well.

3.3. House prices

In UK the HM Land Registry provides data on the prices of properties sold in England and Wales. They allowed us to access data for all properties sold in London. This database has data stored back to 1997, given us easy access for house prices for 2016 to 2017.

The HM Land Registry register property and land in England and Wales worth in excess of £4 trillion, including close to £1 trillion of mortgages. Anyone that is buying or selling property, or taking out a mortgage must apply to the register:

- unregistered property
- any new owner of registered property
- an interest affecting registered property, such as a mortgage, a lease or a right of way

(HM Land Registry, 2018)

With the price of the property sold and type of property (Detached, semi-detached, terrace, flat/mansionette), we were able to control for some of the house characteristics. However, data on sq. meters, room number etc. were not available, so were not able to control for all house characteristics.

Nonetheless, we were able to get the average house price sold for every postcode sector in London.

This is the website that provide the UK House Price Index, and the same information can be accessed on GOV.UK. Therefore, we believe that this is the best source of house price information.

3.4. Neighborhood characteristics

As demographic information we included: % of people under 30, ethnicity, % of couples with or with/without dependent children, % of lone parents, % one-person households and media income. These where chosen, as we believe this was the available data that would have most impact on house prices.

Neighborhood characteristics are taken from the 2011 Census. The 2011 Census was completed on 27 March 2011 by Northern Ireland Statistics & Research Agency (NISRA), National Records of Scotland (NRS), and the Office for National Statistics (ONS). The Census is created

every 10 years to provide a detailed picture of the population and its characteristics in an effort to allows central and local government, health authorities and many other organizations to target their resources more effectively and to plan housing, education, health and transport services for years to come (Office of National Statistics, 2018).

Output Areas (OA) are the lowest geographical area for which the census provides information. In England and Wales there are in total 181,408 OAs. Super Output Areas (SOA) are created to improve the reporting of small area statistics by grouping together OAs with similar demographic information. The two SOAs that are most commonly used are Lower Layer Super Output Area (LSOA) and Middle Layer Super Output Area (MSOA). (ONS, 2018).

LSOAs are created from 4 to 6 OAs with similar demographic information with a mean population of 1500 people (NHS, 2018). Whereas MSOAs are created from a group of LSOAs with similar demographic characteristics, with a mean population of 7000 (NHS, 2018).

As it is a legal requirement to complete the Census survey, we believe that this is a good indicator of neighborhood characteristics. We used the neighborhood characteristics based on MSOAs to match to postcodes. This in turn allowed us to fund out the neighborhood characteristics of the area that surrounds the school, which we later used for matching purposes.

3.5. Crime rate

From the crime data available from the London Datastore, we decided to include *"violence against the person"* which includes homicide, Death or Serious Injury – Unlawful Driving, Violence with injury, Violence without injury and Stalking and Harassment. We also included *"sexual offences"* which includes rape, and different forms of sexual assault (Home Office Counting Rules For Recorded Crime, 2018). Basically, we focused on crimes which would affect people physically.

We were able to collect crime rate data for all the years needed. The data was downloaded from London Datastore, which is a website where it is easy to access data about London. The Metropolitan Police provide updated crime statistics on a monthly basis here. No information was available on MSOA geographical level, so we opted for borough level crime statistics. The problem with borough level crime statistics is that crime levels are not the same over the whole borough, and some borough have both good and bad areas. However, the benefit of boroughs level crime statistics is that crime is often reported in the news at borough level. Meaning that

when a crime is reported in the news, when they describe where it has happened, the borough is often named, not the specific street. For example, Hackney used to be a considered a bad area in London, and the news did often specify where in hackney the high crime rates where, just that Hackney had a high crime rate. As many people get their information on crime from the news, one can assume that people are more concerned by borough level crime statistics.

This gave us 198,011 reported *violence against person offences* in 2017 and 226,618 in 2016. For *sexual offence*s we got 14,539 reported offences in 2017 and 16,237 in 2016.

We chose to include violence against person and sexual offences, as we believe those are the crimes that people are scared off, as they often result in physical injury and this is often what people refer to when they are talking about crime rate. If we only took the total figure for crime, areas which a lot of tourists would have the highest crime rate, because of the amount if pick pocketing. For example, Westminster and Camden, have the highest total recorded offences in London, but are both very desirable places to live. Therefore, only violence against person and sexual offences were included.

3.6. School demographics

We also wanted to include data from the National Pupil Database (NPD), as it would give us a detailed demographic breakdown of every school. However, due to time restrictions and uncertainty on whether we would be able to comply with the Data Protection Act 1998 (UK), we did not include this information in our research.

3.7. Method used for data connection

Finally, in order for us to match the Census information, which is categorized based on MSOA areas, we downloaded additional data which linked MSOA areas to postcodes. This enabled us to match the Census data with house prices and schools, based on their postcodes. Crime rate was matched based on MSOA areas, meaning that all the MSOA areas that were included in a borough would share the same crime rate.

4. Methodology

4.1. Propensity score matching

After Rosenbaum and Rubin ‘s (1983) paper on propensity score analysis, this method has been increasingly popular over the last three decades (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015). This method has been widely used in social, behavioral and medical research, and few in nursing research. (Pan & Bai, Propensity Score Methods in Nursing Research, 2016). We opted for this method as we cannot use school catchment areas to measure difference in house prices, as in Black’s (1999) paper.

4.2. Definitions

Propensity score is the conditional probability of a subject being assigned to a treatment group, given the observer covariates (Pan & Bai, Propensity Score Methods in Nursing Research, 2016). In our model, good performance on schools are in the treatment group, while bad and average schools are in the control group. In other words, housing prices with good schools around are in the treatment group and the rest are in the untreated group.

Propensity score methods are statistical methods using propensity scores to balance the distributions of observable baseline covariates between the treatment and control group, with an aim to reduce selection bias (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015). In other words, the propensity score is a balancing score, which means the distribution of measured baseline covariates is similar between treated and untreated subjects based on the propensity score. Thus, in a set of subjects all of whom have the same propensity score, the distribution of observed baseline covariates will be the same between the treated and untreated subjects (Austin, 2011). This will enable us to do a direct comparison between the treatment and control group (Pan & Bai, Propensity Score Methods in Nursing Research, 2016).

4.3. Treat and Control groups

Prior to the analysis, the data will first be split into treatment and control groups. The evaluation and grouping of the data based on school performance will be using two difference score methods - the percentage of students that meet the expected standard and the progress scores.

When we look at the "% of students that meet expected standard in math, reading and writing" as an indicator of school quality, school test scores are first sorted from highest to lowest for both 2016 and 2017. Schools are then picked and grouped as good performing schools if they rank in top quartile in both 2016 and 2017. This gives us 128 schools with s consistently high scores out of 813 public schools in the dataset. We have done it this way as one can assume that school reputation comes from a school consistently performing well. However, we are aware that two years are not sufficient to assume that a top performing school has a good reputation, but it is better than just considering one year. If we had an option of more years we would opt for that.

For "progress scores", school performance is split into progress scores on reading, writing and math. The data covers all three sections for continuous two years, 2016 and 2017, which gives us 6 columns of test scores. We decide to choose schools with scores above zero (above average) in at least 5 columns. That leaves us 116 good performing schools out of 813 in the dataset.

To observe how school quality affects nearby property market price, we analyze our data set and describe property sales price as regression of observable property characteristics. As mentioned in the literature review, it is not possible to observe all property characteristics that are significant towards affecting property sales price. Hence, implementing a regression directly can lead to omitted variable bias. This is a limitation when it comes to measuring school performance accurately as it creates an endogeneity problem.

For example, the best schools in London are more likely to be situated in better neighborhoods. However, another view point is that wealthy neighborhoods tend to attract families with better socio-economic backgrounds and children from these families tend to perform better at school as compared to their peers from lower socio-economic communities. This implies that house prices are not solely driven by school quality and can be affected by other desired neighborhood qualities – better employment opportunities, popular shopping malls, and neighborhood peers, etc. On top of it being difficult to isolate the effect of schools on house prices from the impact of other traits, there also exist endogeneity between variables.

4.4. Technical explanation for propensity score matching

If we have **N** units (subjects), **z** will be the treatment condition and r will be the potential response. For every unit **i** (i=1, …, N), $\mathbf{z_1 = 1}$ indicated that the unit **i** in the treatment group with the potential corresponding response $\mathbf{r_{1i}}$, and $\mathbf{z_1 = 0}$ showing that unit **i** in the control group with a potential corresponding result of $\mathbf{r_{0i}}$. What we are interested in is the treatment effect of each unit **i,** $\Delta_i = \boldsymbol{r_{1i}} - \boldsymbol{r_{0i}}$. However, as every unit **i** cannot be in the treatment and control group at the same time, a way to estimate this would be the *average treatment effect* (ATE). This is defined as $ATE = E(r_1 - r_0) = E(r_1) - E(r_0)$, where we will find $E(r_1)$ = the expected value of r for all units in the treatment group and $E(r_0)$ = the expected value of r for all the units in the control group. When looking at *randomized control groups* (RTC) the ATE will be an unbiased estimate as the treatment group does not on average differ too much from the control group, based on their observable and unobservable characteristics, due to randomization (Rubin, 1974)**.** In non-RCTs, ATE could be biased as the treatment and control groups may not be comparable, due to group selection bias in the observable data. Selection bias which can be overt, hidden or both (Rosenbaum P. M., 2010), but propensity score analysis can reduce overt bias in observable studies, by balancing the distribution of covariates between the treatment and control group. Therefore, we can get an unbiased estimate of the ATE using propensity score analysis using observable data. (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015)

However, we are trying the find the *average treatment effect for the treated* (ATT), which is defined as $ATT = E(r_1 - r_0 | z = 1) = E(r_1 | z = 1) - E(r_0 | z = 1)$. This is because we are trying to find the effect a good school (z=1) will have on surrounding house prices. Here we will have the same problem, as r_0 can't be observed when $z = 1$. To get around this problem we will use propensity score matching, by matching units in the treatment group with units in the control group based on propensity scores. Since the $z = 1$ have similar covariates compared to the control group, one can use propensity score matching to estimate ATT. (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015)

Our propensity score method consist of three steps:

1. Selecting covariates;
2. Estimating propensity scores;

3. Matching covariates using propensity scores.

4.5. Propensity score matching steps

4.5.1. Selecting covariates

Selecting covariates was done by looking at what previous papers deemed important and what we intuitively believed to be important to parents when they make a decision to buy a house.

4.5.2. Using Propensity Score Estimation

The propensity score for a unit i, $e(X_i)$, has been estimated by using a logistic regression of the treatment condition z_i, on the vector of covariate X_i (Agrestic, 2013):

$$\text{Equation 1: P (T} = 1 | X_1, \dots, X_\text{i}) = \exp(X_i) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_\text{i} X_i)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_\text{i} X_\text{i})}$$

(Grotta & Bellocco, 2014)

In the above equation β is a vector of regressed coefficients. The log of the propensity scores is often used. (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015)

In the regression model that is used in the propensity score method, we first introduce dummy variables to represent good-performing schools and bad-performing schools. The regression is described as:

$$\text{Equation 2: } price_{jk} = \alpha + \beta_1 \Phi_{score} + \beta_2 H'_{jk} + \beta_3 N'_k + \varepsilon_{jk}$$

$$where\ price_{jk} = property\ sales\ price\ in\ 2017, in\ MSOA\ j, near\ school\ k$$

$$\Phi_{score} = 0\ or\ 1, and\ denotes\ test\ score\ dummies$$

$$\beta_1 = coeffiecient\ of\ good\ test\ scores$$

$$H_{jk} = \text{vector of the observed characteristics of houses in MSOA } j$$

$$N_k = \text{vector of observed neighbourhood and school district characteristics}$$

$$\varepsilon_{ijk} = error\ term$$

Note that H_{jk} and N_k is represented by $X_1 + \cdots + X_i$ in equation 1

4.5.3. Matching covariates using propensity scores

There are a few ways to match propensity scores. The most basic method is called *nearest neighbor* and it matches every unit **i** in the treatment group with a unit **j** in the control group that has the closest propensity score, defined as: $d(i,j) = min_j\{|e(X_i) - e(X_i)|\}$.

Another way is to use *radius matching* where you match unit **i** in the treatment group to a number of units in the control group within a determined band **b**; defined as $d(i,j) = \{|e(X_i) - e(X_i)| < b\}$. There are also matching methods like stratification that do not match individual units, which classifies all units in the sample into a number of strata based on the number of percentiles. It has been argued that using this method and classifying the dataset into five strata removes close to 90% of selection bias. Finally, kernel matching in which the treated units are matches on a weighted average of all the control group. The closer the control is to the treated the higher weight it will get**.** (Pan & Bai, Propensity Score Analysis: Fundamentals and Developments, 2015)

Even though selecting the right matching method is important, it has been argued that it is more important to select the correct covariates (Steiner & Cook, 2013).

To match up properties based on observed identical or similar neighborhood and house characteristics, we pair up each house in the treated group with a corresponding house in the untreated group on the basis of identical or similar characteristics such as demographics and house types. We implement two different matching methods – the nearest neighbor matching and Kernel matching to see outcomes, as this gives us a measure of one-on-one and one-to-all measure. Finally, we can calculate the impact of good school performance on housing prices by comparing the means of outcomes across treated participants and their matched pairs. Since having ensured that houses in two different groups share the same or similar demographics and characteristics, test score differences would be the only reason for housing price differences.

5. Descriptive Statistics

The entire sample covers 813 public primary schools where students between the ages of 7 to 11 years in 593 different MSOA neighborhoods in the Greater London areas.

Table 5.1 is the summary of key statistics of the entire data. The mean sales price is £549,996.60 with a standard deviation of £30,4211.90 in the year of 2017.

We use 2011 Census demographics information as a proxy for characteristics of neighborhoods. When moving to a new place, crime rate in the neighborhood is ranked as the first thing that people consider (Unpaktblog, 2018). To satisfy the balancing property, the log of number of crime cases is used in the model. The mean log number of violence against person cases and sex offence cases that happened in 2017 are 8.75 and 5.95, respectively per MOSA area. The average age structure is 43% for people who are under the age of 30, and 46% for people at the age above 30 but below 65. Previous research has suggested that ethnicity is important to parents when choosing schools, so we have included ethnic demographic information as well. Mean of the log percentage composition of white, Asian, black and mix ethnicity in MSOA are -0.63, -1.94, -2.24 and -3.05 respectively. Households of couples with dependent children and without dependent children are each approximately 18%. Whereas, 13% are lone parents and 30% are households of one person. House characteristics are also taken into consideration in our model, where 6% of properties are detached; 19% are semi-detached; 25% are terraced. Amongst the different housing types, most live in a flat or apartment, taking up 49% of the total population.

Variable	Obs	Mean	Std. Dev.	Min	Max
treat	813	.1574416	.3644408	0	1
avepri2017	813	549996.6	304211.9	239300	4250073
lnvio2017	813	8.748556	.3040749	7.890583	9.176
lnsexoff2017	813	5.952393	.5753091	3.135494	6.6
ageunder30	813	.4327454	.0579775	.297	.591
ageover30~65	813	.4603862	.0364887	.34	.562
logwhite	813	-.6345633	.4422087	-2.797	-.039

logmix	813	-3.054048	.3689474	-4.51	-2.313
logasian	813	-1.944838	.7798273	-4.075	-.207
logblack	813	-2.23599	.8624261	-4.962	-.667
couplewith~n	813	.1832964	.0471596	.07	.322
couplenode~n	813	.1814637	.0441789	.087	.303
loneparent	813	.1367196	.0480479	.036	.28
oneperson	813	.3003173	.0617576	.126	.546
detached	813	.0610467	.0772274	.003	.653
semidetached	813	.19031	.1580801	.005	.889
terraced	813	.2550738	.1458967	.006	.683
flatorapar~t	813	.4927134	.2470571	.022	.986
medianincome	813	33546.82	9840.855	17877.69	71771.21

Table 5.1: Summary of treatment, X variables and Y variables

5.1. Score method 1: Percentage of students that meet the expected standard

Table 5.2 below illustrates the results when we look at the "% of students that meet the expected standard in reading, writing and math" as an indicator of school quality. It presents the data summary for treated group and untreated group separately. With 128 observations in the "treatment" group, the mean sales price is £598,851.90 (standard deviation of £292,525.70). The lowest price is £275,455 and the largest one is £2,319,214. The sample of the untreated group consists of 685 primary schools with an average associated house price of £540,867.40 (standard deviation of £305,687.60). The lowest price is £239,300 and the largest one is £4,250,073.

treat = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
avepri2017	685	540867.4	305687.6	239300	4250073
lnvio2017	685	8.755188	.3026589	7.890583	9.175542

lnsexoff2017	685	5.968416	.5572854	3.135494	6.597146
ageunder30	685	.4347197	.0571862	.297	.591
ageover30~65	685	.4590336	.0363533	.34	.562
logwhite	685	-.6490964	.446065	-2.797	-.039
logmix	685	-3.047028	.3621059	-4.51	-2.313
logasian	685	-1.929407	.7780548	-4.075	-.207
logblack	685	-2.199469	.8458055	-4.962	-.667
couplewith~n	685	.1834277	.0463166	.074	.322
couplenode~n	685	.1798409	.0431504	.087	.303
loneparent	685	.1390934	.0476765	.039	.28
oneperson	685	.2987416	.060228	.146	.546
detached	685	.0584555	.0705015	.003	.653
semidetached	685	.1924832	.1569646	.005	.889
terraced	685	.2573182	.1468676	.006	.683
flatorapar~t	685	.4909314	.2449843	.022	.986
medianincome	685	32812.7	9295.897	17877.69	71771.21

treat = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
avepri2017	128	598851.9	292525.7	275455	2319214
lnvio2017	128	8.713063	.3103382	7.891	9.176
lnsexoff2017	128	5.866641	.6593438	3.14	6.6
ageunder30	128	.4221797	.0611976	.315	.546
ageover30~65	128	.467625	.0365011	.391	.554

logwhite	128	-.5567891	.4140095	-2.343	-.07
logmix	128	-3.091617	.4030422	-4.51	-2.397
logasian	128	-2.027414	.7871685	-3.612	-.265
logblack	128	-2.431437	.9256642	-4.711	-.738
couplewith~n	128	.1825938	.0516267	.07	.305
couplenode~n	128	.1901484	.0485867	.087	.292
loneparent	128	.1240156	.0482198	.036	.264
oneperson	128	.30875	.0690211	.126	.493
detached	128	.0749141	.1055102	.004	.524
semidetached	128	.1786797	.1640636	.006	.667
terraced	128	.2430625	.1405342	.024	.551
flatorapar~t	128	.50225	.2586539	.048	.952
medianincome	128	37475.51	11632.25	19212.34	67499.91

Table 5.2: Summary of X variables and Y variable sort by treatment on expected scores

5.2. Score method 2: Progress scores

The table 5.3 below illustrates the results when we look at "progression scores" as an indicator of school quality. With 116 observations in the "treatment" group, the mean sales price is £572,674.40 with a standard deviation of £273,988.80. The lowest price is £275,455 and the largest one is £2,319,214. The sample of the untreated group consists of 697 schools with an average of £546,222.40 in housing price and a standard deviation of £308,974.60. The lowest price is £239,300 and the largest one is £ 4,250,073.

treat = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
avepri2017	697	546222.4	308974.6	239300	4250073
lnvio2017	697	8.739465	.3084636	7.890583	9.175542

lnsexoff2017	697	5.951821	.6026644	3.135494	6.597146
ageunder30	697	.4307532	.0574925	.297	.591
ageover30~65	697	.4602597	.0359272	.34	.562
logwhite	697	-.6285409	.4361788	-2.797	-.039
logmix	697	-3.064584	.3774075	-4.51	-2.313
logasian	697	-1.947298	.780622	-4.075	-.207
logblack	697	-2.256953	.8784855	-4.962	-.667
couplewith~n	697	.1860846	.0466567	.074	.322
couplenode~n	697	.1828881	.0445698	.087	.303
loneparent	697	.1370215	.0482784	.036	.28
oneperson	697	.2973515	.0600627	.126	.546
detached	697	.0634175	.0796386	.003	.653
semidetached	697	.2012152	.1613668	.005	.889
terraced	697	.2571521	.1467692	.006	.683
flatorapar~t	697	.477373	.2450638	.022	.986
medianincome	697	33483.32	9756.079	17877.69	71771.21

treat = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
avepri2017	116	572674.4	273988.8	275455	2319214
lnvio2017	116	8.803167	.2710879	7.890583	9.175542
lnsexoff2017	116	5.956015	.374231	5.393628	6.597146
ageunder30	116	.4447155	.059674	.318	.576
ageover30~65	116	.4611466	.039855	.365	.554

logwhite	116	-.67075	.4772469	-2.797	-.093
logmix	116	-2.990741	.3072059	-4.075	-2.442
logasian	116	-1.930052	.7782393	-3.576	-.207
logblack	116	-2.110034	.7499155	-4.51	-.803
couplewith~n	116	.1665431	.0468898	.07	.272
couplenode~n	116	.1729052	.0408975	.092	.27
loneparent	116	.1349052	.0468013	.04	.28
oneperson	116	.3181379	.0687251	.176	.493
detached	116	.0468017	.0590599	.004	.427
semidetached	116	.1247845	.1174753	.006	.566
terraced	116	.2425862	.1405062	.024	.551
flatorapar~t	116	.5848879	.2398091	.096	.952
medianincome	116	33928.34	10371.5	19212.34	67499.91

Table 5.3: Summary of X variables and Y variable sort by treatment on progress scores

6. Results

6.1. Regression with a dummy variable

6.1.1. Score method 1: For percentage of students that meet the expected standard

Table 6.1 below is the results of the regression with dummy variables for treatment. The difference is £57,984.50, meaning the "treatment" increases the housing price by that amount as a direct comparison is done between the outcomes of the two groups. This is the outcome when we do not control for other factors.

avepri2017	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
treat	57984.5	29240.75	1.98	0.048	588.0298	115381
_cons	540867.4	11602.41	46.62	0.000	518093.1	563641.7

Table 6.1: Regression with a dummy variable for treatment (t-test) on percentage scores

The Table 6.2 shows the results for equation (2), which controls for other independent variables, while taking account into demographics and house characteristics. For example, crimes including violence against person and sex offense are considered; age distribution such as the percent of age under 30 and age between 30 and 65, and ethnicity groups such as white, Asian, black and mix are included. Household types consisting of couples with or without dependent children, lone parent, one person, and house types consisting detached, semi-detached, terraced, and flat or apartment are also controlled. Finally, the annual median income is also taken into consideration. When we control for independent variables, we get a different outcome from the previous regression. The mean property price in the treated group is £5,877.71 more than that in the untreated group.

avepri2017	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
treat	5877.707	24381.11	0.24	0.810	-41981.35	53736.77
lnvio2017	-46997.19	37135.5	-1.27	0.206	-119892.5	25898.17
lnsexoff2017	26457.12	16769.62	1.58	0.115	-6460.915	59375.16

ageunder30	-1666997	486000.5	-3.43	0.001	-2620995	-712999.2
ageover30under65	214869.7	622109.6	0.35	0.730	-1006304	1436044
logwhite	35254.91	49758.67	0.71	0.479	-62419.19	132929
logmix	186690.6	49377.74	3.78	0.000	89764.21	283616.9
logasian	-28481.02	22536.44	-1.26	0.207	-72719.07	15757.02
logblack	-133266.1	26634.52	-5.00	0.000	-185548.4	-80983.66
couplewithdependent children	-1002223	399447.9	-2.51	0.012	-1786322	-218124.6
couplenodependentchildren	-2781288	609034.7	-4.57	0.000	-3976796	-1585779
loneparent	-1101358	459641.7	-2.40	0.017	-2003614	-199101
oneperson	-1324365	399811.4	-3.31	0.001	-2109178	-539553.2
detached	2650814	3145977	0.84	0.400	-3524602	8826229
semidetached	2735528	3117207	0.88	0.380	-3383413	8854469
terraced	2849666	3111657	0.92	0.360	-3258381	8957713
flatorapartment	3287776	3110698	1.06	0.291	-2818388	9393940
medianincome	5.030916	2.043445	2.46	0.014	1.019722	9.04211
_cons	-296917.2	3237589	-0.09	0.927	-6652163	6058328

Table 6.2: Regression with a dummy variable for treatment controlling for x on percentage scores

6.1.2. Score method 2: Progression scores

We replicate the same steps as above on “progression scores”.

Table 6.3 presents the results of the regression with dummy variables for treatment. The mean property sales price in treated group is £26,452.01 higher than it in the untreated group when we do not consider for house characteristics.

avepri2017	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]

treat	26452.01	30510.02	0.87	0.386	-33435.92	86339.94
_cons	546222.4	11524.61	47.40	0.000	523600.8	568844

Table 6.3: Regression with a dummy variable for treatment (t-test) on progression scores

When we control for independent variables in Table 6.4, the difference between mean property price in the treated group and in the untreated group is -£24,131.18. Properties with high quality schools nearby have lower mean housing prices than the ones with average and bad schools around. However, we still need to eliminate the bias problem caused by omitted variables. Hence, we introduce the propensity score matching method to fix it.

avepri2017	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
treat	-24131.18	25192.08	-0.96	0.338	-73582.13	25319.76
lnvio2017	-47738.73	37053.25	-1.29	0.198	-120472.6	24995.17
lnsexoff2017	26203.67	16732.4	1.57	0.118	-6641.293	59048.63
ageunder30	-1654304	485829.3	-3.41	0.001	-2607965	-700642.3
ageover30under65	243046.4	621974.4	0.39	0.696	-977862	1463955
logwhite	36039.19	49735.87	0.72	0.469	-61590.14	133668.5
logmix	183664.2	49224.1	3.73	0.000	87039.49	280289
logasian	-28021.28	22510.08	-1.24	0.214	-72207.58	16165.03
logblack	-132678.8	26598.66	-4.99	0.000	-184890.8	-80466.83
couplewithdependentchildren	-1024739	399816.1	-2.56	0.011	-1809561	-239917.7
couplenodependentchildren	-2815498	609120.3	-4.62	0.000	-4011174	-1619821
loneparent	-1106957	459166.7	-2.41	0.016	-2008281	-205632.9
oneperson	-1323945	399182.9	-3.32	0.001	-2107523	-540366.3
detached	2589535	3144121	0.82	0.410	-3582236	8761307
semidetached	2666598	3115275	0.86	0.392	-3448551	8781746
terraced	2786197	3109706	0.90	0.371	-3318020	8890414
flatorapartment	3222746	3108706	1.04	0.300	-2879508	9325000
medianincome	5.178709	2.023636	2.56	0.011	1.206399	9.151019
_cons	-238450.4	3235579	-0.07	0.941	-6589750	6112849

Table 6.4: Regression with a dummy variable for treatment controlling for x on progression scores

6.2. Propensity score matching

6.2.1. Propensity score

We implement a propensity score matching model by matching the propensity scores. According to output, for score "% of students that meet the expected standard in reading, writing and math:, the range of propensity score is 0.049 to 0.634, and the number of blocks is 5 in our model; for "progression score", the range of propensity score is 0.034 to 0.314, and the number of blocks is 3. Within these blocks, we would have observations with similar characteristics.

6.2.2. Nearest Neighbor Matching Method

Using the nearest neighbor matching method based on "% of students that meet the expected standard in reading, writing and math", we get the results shown in Table 6.5 below. The number of treated observations is 128 and we have found 124 observations in the control group as their nearest neighbors, which means we cannot find a unique matching partners for all treated ones. The difference of £50,442.68 between the outcomes of treatment and control group after matching up is the effect of the "treatment". In other words, if the "% of students that meet the expected standard in reading, writing and math" is ranked top 25 percentile among all the schools across the Greater London area for two continuous years 2016 and 2017, then the surrounding properties have on average £50,442.68 higher value.

n. treat.	n. contr.	ATT	Std. Err.	t
128	124	50442.680	42585.568	1.185

Table 6.5: ATT estimation with Nearest Neighbor Matching method on percentage scores

Table 6.6 describes that based on "progression scores", the mean house price of the control group is £57,000 higher than that of the treatment group. This is interesting as we get a lower average value for houses around schools with an above average progression score.

n. treat.	n. contr.	ATT	Std. Err.	t

116	120	-5.70e+04	52267.021	-1.090

Table 6.6: ATT estimation with Nearest Neighbor Matching method on progression scores

6.2.3. Kernel Matching Method

We decided to use Kernel Matching Method as well, to see if this would give us a different outcome.

In Table 6.7 it shows "% of students that meet the expected standard in reading, writing and math", number of treated observations is 128, and each of them is matched with at least one unit in the control group. The total number of untreated observations is 682. The difference between the outcomes of the treatment and the control groups is £22,408.23 with a standard error of £28,099.10.

n. treat.	n. contr.	ATT	Std. Err.	t
128	682	22408.23	28099.10	0.797

Table 6.7: ATT estimation with the Kernel Matching method on percentage scores

Table 6.8 presents the results for "progression scores", of 116 schools in the treatment group with a higher mean house price of £3,406.28 (with a standard error of £28,099.10).

n. treat.	n. contr.	ATT	Std. Err.	T
116	667	3406.276	28609.936	0.119

Table 6.8: ATT estimation with the Kernel Matching method on progression scores

With the "% of students that meet the expected standard in reading, writing and math" we reach the same conclusion using both matching methods, that school quality has a positive influence on the housing price of surrounding properties. Whereas for "progression scores" we get a negative value for nearest neighbor and a positive value for Kernel. It is important to note that 58 schools that are considered good schools when we look at progression scores are not good schools if we look at "% of students that meet the expected standard in reading, writing and math".

7. Conclusion

Based on our results we can see that a good school will most likely increase the average house price of the surrounding area. This would be something that one could intuitively assume to be true. But what we found interesting was that "progression scores", which the UK government have been pushing for the last two years, do not necessarily have a positive effect on surrounding house prices. "*% of students that meet the expected standard in reading, writing and math"* on the other hand seems to have a positive effect. Why is this?

We believe that this is because "*% of students that meet the expected standard in reading, writing and math"* is a metric that is easy to understand for a parent that is looking at what is the best school. This is because the assessment process is more or less just looking at which school had the most students meet the required standards as a school that has many students with high grades can be assumed to be a good school. However as mentioned earlier, this can also be a result of improved enrolment quality or greater pupils' progress, not necessary better school performance. Similar measures as this is probably what most parents have looked at historically, therefore we believe that this score would result in a positive effect.

On the other hand, with progression score we got both a positive and negative result. This is interesting, as we believe this could be a result of the fact that this is a new score, harder score to understand and not all schools are compared to each other. Schools with low scoring KS1 students are compared to schools with similar KS1 scoring students. Meaning that a school could have really good scoring KS1 students, but because they then did progress to the same extent as the school with bad KS1 scoring student, in our study they were not considered a good school (when it comes to progression scores), even though their KS2 scores where higher.

We believe that progression scores are a better indicator of school quality, as it basically measures the improvement of students throughout primary school. Nonetheless, we believe that the "% of students that meet the expected standard in reading, writing and math" will have more of an impact on price, as it is easier to intuitively understand. If more data past yearly data was available for progression score, we believe that we would be able to come up with a more definite answer.

When looking that the different matching methods, we believe that Kernel matching is probably more accurate as it uses a weighted average, whereas the nearest neighbor method tries to find

the closes match. Since nearest neighbor is more basic compared to Kernel, we believe that the Kernel result is more accurate.

Our method only shows how much house prices will be higher in an area on average, but it would be good to know how much house prices increase for different levels of school quality. Moreover, more detailed information on house characteristics such as sq. meters, number of rooms etc., would make the result even more accurate. Another point to take note is that our data set focuses on state funded schools and does not take into account private schools in London. The influence of school quality for private schools can be expanded on from our study, and the influence could possibly be more significant as wealthier families are more likely to be willing to pay higher premiums in order to secure housing near the private school of their choice.

To conclude we believe that good schools will increase the price of surrounding house prices by an average of £22,408.23 or 3.74%. This is similar to previous research, even though they used a different method. However, it would be interesting to see what the effect the "progression score" will have on house prices over the coming years.

8. References

Austin, P. C. (2011, May). *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies*. Retrieved from www.ncbi.nlm.nih.go: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483/#

Bellocco, A. G. (2104, November 13). *A review of propensity score: principles, methods and application in Stata*. Retrieved from https://www.stata.com/meeting/italy14/abstracts/materials/it14_grotta.pdf

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics Vol.114*, 577 - 599.

Black, S., & Machin, S. (2009). Housing Valuation of School Performance. *Handbook of Economics of Education*.

Black, S., & Machin, S. (2010). Housing Valuations of School Performance. *Handbook of the Economics of Education*.

CEGA UC Berkeley. (n.d.). *Propensity-Score Matching (PSM).* Retrieved from cega.berkeley.edu/assets/cega_events/31/Matching_Methods.ppt

Clapp, J. M., Nanda, A., & Ross, S. L. (2007). Which School Attributes Matter? The Influence of School District Performance and Demographic Composition on Property Values. *Economics Working Papers*, http://digitalcommons.uconn.edu/econ_wpapers/200526.

DfE. (2016). *GOV.UK*. Retrieved from Primary Progress measure: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/560969/Primary_school_accountability_summary.pdf.pdf

Fack, G., & Grenet, J. (2008). When do Better Schools Raise Housing Prices? Evidence from Paris Public and Private Schools.

Figlio, D. N., & Lucas, M. E. (2004). What's in a Grade? School Report Cards and the Housing Market. *The American Economic Review Vol. 94, No. 3*, 591-604.

Gibbons, S., Machin, S., & Silva, O. (2012, October). *Valuing school quality using boundary discontinuity*. Retrieved from LSE Research Online: http://eprints.lse.ac.uk/45246/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_sh

ared_repository_Content_Silva%2C%20O_Valuing%20school%20quality_Silva_Valuing%20school%20quality%20_2014.pdf

GOV.uk. (2018, 05 10). *Type of Schools*. Retrieved from GOV.UK: https://www.gov.uk/types-of-school

Grotta, A., & Bellocco, R. (2014, 11 13). *A review of propensity score: principles, methods and application in Stata.* Retrieved from https://www.stata.com/meeting/italy14/abstracts/materials/it14_grotta.pdf

Hackney Council. (2017). *Hackney's Primary Schools 2018 Admission Guide.* Retrieved from Hackney Learning Trust: https://www.learningtrust.co.uk/sites/default/files/document/Primary%20Schools%20Admissions%20Guide.pdf

HM Land Registry. (2018). *About us*. Retrieved from GOV.UK: https://www.gov.uk/government/organisations/land-registry/about

Home Office Counting Rules For Recorded Crime. (2018, April). *GOV.UK.* Retrieved from Guidance Counting rules for recorded crime: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662135/count-offence-classification-index-nov-2017.pdf

LondonDatastore. (2018, 05 23). *London's current population is 8.904 million.* Retrieved from London Datatore: https://data.london.gov.uk/demography/

Machin. (2011). Houses and Schools - Valuation of School Quality Through the Housing Market. *Labour Economics, 18*, 723-29.

Machin, S., & McNally, S. (2008). The Literacy Hour. *Journal of Public Economics, vol. 92*, 1441-1462.

NHS. (2018). *Lower Layer Super Output Area* . Retrieved from NHS Business Definitions: https://www.datadictionary.nhs.uk/data_dictionary/nhs_business_definitions/l/lower_layer_super_output_area_de.asp?shownav=1

NHS. (2018). *Middle Layer Super Output Area*. Retrieved from NHS Business Definitions: https://www.datadictionary.nhs.uk/data_dictionary/nhs_business_definitions/m/middle_layer_super_output_area_de.asp?shownav=1

OECD. (2015, 05). *What do Parents look for in their chiled's school?* Retrieved from The Organisation for Economic Co-operation and Development: https://www.oecd.org/pisa/pisaproducts/pisainfocus/PIF-51(eng)-FINAL.pdf

Office of National Statistics. (2018). *What is a census?* Retrieved from ONS: https://www.ons.gov.uk/census/2011census/howourcensusworks/aboutcensuses/whatisacensus

Ofsted. (2018, 05 23). *GOV.UK*. Retrieved from About us: https://www.gov.uk/government/organisations/ofsted/about

ONS. (2018). *Census geography*. Retrieved from ONS.gov.uk: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#output-area-oa

ONS. (2018). *Overview of violent crime and sexual offences*. Retrieved from Office of National Statistics: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/compendium/focusonviolentcrimeandsexualoffences/yearendingmarch2016/overviewofviolentcrimeandsexualoffences

Pan, W., & Bai, H. (2015). *Propensity Score Analysis: Fundamentals and Developments.* New York: Guilford Press; 1 edition.

Pan, W., & Bai, H. (2016, August 31). Propensity Score Methods in Nursing Research. Retrieved from Nursing Research: file:///C:/Users/nsampson/Downloads/Propensity_Score_Methods_in_Nursing_Research__Take.1.pdf

Rosenbaum, P. M. (2010). *Observable studies (2nd edition).* New York: Springer-Verlag.

Rosenbaum, P. R., & Rubin, D. B. (1983, April). *The Central Role of the Propensity Score in Observational Studies for Causal.* Biometrika Trust. Retrieved from http://links.jstor.org/sici?sici=0006-3444%28198304%2970%3A1%3C41%3ATCROTP%3E2.0.CO%3B2-Q

Rubin, D. B. (1974). Estimating causal effect of treatment in randomized adn nonrandomized studies. *Journal of Educational Psycology 66(5)*, 688-701.

Schneider, M., & Buckley, J. (2002). What Do Parents Want from Schools? Evidence from the Internet. *Educational Evaluation and Policy Analysis Vol. 24, No. 2*, 133-144.

Standar & Testing Agency. (2017). *Information for schools and parents: Primary school progress measures.* Retrieved from GOV.co.uk: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/596097/STA177916e_Primary_school_progress_measures_leaflet_V6_FINAL.pdf

Standards and Testing Agency. (2017, July 4). *GOV.UK*. Retrieved from Scaled scores at key stage 2: https://www.gov.uk/guidance/scaled-scores-at-key-stage-2

Steiner, P. M., & Cook, D. (2013). Matching and Propensity scores. *The Oxford handbook of quantitative methods (Vol 1)*, 237-259.

Stephen Gibbons, S. M. (2012, October). *Valuing school quality using boundary discontinuity.* Retrieved from LSE Research Online: http://eprints.lse.ac.uk/45246/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_Silva%2C%20O_Valuing%20school%20quality_Silva_Valuing%20school%20quality%20_2014.pdf

The Organisation for Economic Co-operation and Development. (2015, 05). *What do Parents look for in their chiled's school?* Retrieved from OECD: https://www.oecd.org/pisa/pisaproducts/pisainfocus/PIF-51(eng)-FINAL.pdf

Unpaktblog. (2018, 5 21). *Things to Consider When Moving to a New Neighborhood / City*. Retrieved from Unpaktblog: http://www.unpakt.com/blog/things-consider-moving-new-neighborhood-city/

Wespieser. (2018). *How do parents choose a school for their child?* Retrieved from National Foundation for Educational Research: https://www.nfer.ac.uk/pdf/how-do-parents-choose-school-htu.pdf

Yinger, J., & Nguyen-Hoang, P. (2011). *The Capitalization of School Quality into House Values: A Review.* Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1895351