

# The Impact of School Quality on Property Prices

Joseph Wild

This dissertation applies a hedonic house price model to the city of Sheffield, England to assess the impact of school quality on property prices. Best practice methodology typically applied in US studies has been applied to UK data, using the English Indices of Multiple Deprivation (IMD) to control for neighbourhood characteristics at a highly disaggregated level. For the first time in the UK, actual school catchment areas are used, distinguishing this dissertation from the existing literature, which typically applies a 'straight line distance to the nearest school' approach. A robust OLS regression model is built sequentially to control for property and neighbourhood characteristics, before isolating the impact of school quality. The findings show that houses in catchment areas with 10% more children achieving 5A\*-C at GCSE level can command 7.85% higher house prices. A review of the literature is presented and the rationale for the chosen methodology and variables is discussed, followed by a presentation of the results, which were found to be highly robust.

## 1. Introduction

For many, the assessment of the quality of local schools is an important aspect of the house buying process. A non-economist would likely be able to assert that houses in the catchment areas for good schools will cost more, whilst houses in the catchment areas for poorly performing schools will cost less. But would such a statement actually be correct, and if so, how large would this 'school effect' be? It is this question which the present study seeks to address by utilising a unique dataset and well tested methodology to formalise a topic familiar to so many, but for which a general consensus within academia seems to be lacking.

In the UK much media attention is often given to the schools' admissions process. Phrases found in popular media such as "Choice... is for those who can afford it" (Okolosie, 2016), "£45k to get nice school" (Jones, 2015), "School admissions: the top scams" (Paton, 2009) and "even tougher competition for the most sought-after schools" (Coughlan, 2016) are typical of stories reported by popular newspapers and broadcasters. The media furore is likely to be both a consequence and a cause of the seemingly ever growing public dissatisfaction at the schools' admission process. Many parents are now aware that securing a place at the best state schools requires being resident in the designated catchment areas. The impact of local schools on house prices, is then, seemingly more poignant and more important than ever before, yet remains an under-researched field in the academic sphere.

This dissertation applies a hedonic house price model to assess revealed preferences for education in the housing market. The hypothesis that individuals purchase property with consideration of the quality of local services is by no means new. Indeed, it was first put forward formally by Tiebout as early as 1956. A hedonic house model breaks down the actual transaction price of a property into its component constituents, to model the implicit price of an additional bedroom for example. The theory can be extended to calculate the impact on price of a host of property and neighbourhood features. This research will isolate the impact of education, seeking to assess how much of a property's value is attributable to the quality of local schooling.

To the best of the author's knowledge, this study is the first to provide an estimate of the impact of local education quality on property values using a hedonic house price model in the city of Sheffield. This will allow empirical estimates of the capitalisation of local school quality in property prices for the first time in England's fourth largest city (ONS, 2015). Applying a market valuation technique to assess residents' revealed preferences for education in a city of over 560,000 inhabitants is a notable contribution to the literature. In addition to modelling the hedonic price of education, it

is hoped that the robust regression model developed in this dissertation can be utilised in further research to continue to analyse the impact of *actual* neighbourhood features on property prices.

Reviewing the existing literature, we will find that whilst methodology has advanced in US applications, few studies apply the theoretical developments to UK settings. Section 2 will provide context by outlining the history of hedonic house price analysis and highlight important developments in the literature. Discussion will focus not only on methodology, but also on the critical importance of data selection. It will be noted that models perform better when data is highly geographically disaggregated, as information is lost when properties are grouped into larger geographical areas.

Existing studies are typically plagued by the problem of being unable to successfully isolate the impact of school performance whilst fully controlling for other neighbourhood characteristics. This dissertation overcomes this problem by manipulating data from the English Indices of Multiple Deprivation to directly control for neighbourhood factors. Section 3 will explain how this unique approach represents the forefront of theoretical developments in this field, by utilising data at the most disaggregate level available to explicitly control for local area characteristics. In addition, an explanation for the choice of GCSE results as a measure of school performance will be offered, despite the majority of studies failing to explain the rationale for their chosen indicator.

Section 4 will outline how statistical best-practice is employed to determine the favoured model specification, before results are presented. Results will then be analysed, finding that a one percentage point increase in the proportion of pupils achieving 5A\*-Cs at GCSE level will increase house prices in the local school catchment area by 0.785 percent. Coefficients of other variables will also be discussed, noting the model's implications for factors other than the variable of interest. Section 4 will conclude by employing a host of sensitivity checks to determine that results are indeed robust and are thus reported with a high degree of confidence. Section 5 summarises and concludes the dissertation.

## **2. Literature Review**

This literature review will first explore the origins of the use of hedonic modelling as a technique to decompose house prices into implicit component factors. It will then consider developments in the literature and how methodology has progressed over the past half century. Two notable studies which model the impact of education on property values in the UK context will then be considered. It is found that a gap in the literature exists in the sense that best practice methodology pioneered in US studies has not been applied to UK data. Lastly, the most widely

cited hedonic house price model of the City of Sheffield will be discussed, to identify if any specific locational factors ought to be considered in this study.

### *2.1 Background Theory*

A seminal paper by Wallace Oates (1969) pioneered the use of regression analysis to assess the impact of local public services on house prices. Oates tests Tiebout's (1956) hypothesis that individuals consider the quality of local public services when purchasing property. Tiebout asserts that the market reveals preferences for local public goods, where the provision (or lack) of public services is capitalised into local house prices. Oates finds a positive relationship between house prices and per-pupil school expenditure.

Oates' widely cited study was revolutionary in the sense of using regression analysis to test for the capitalisation of local public services. It was, however, not without criticism. Edel and Sclar (1974) highlight what they claim to be a fundamental theoretical flaw in Oates' work. They argue that the capitalisation of differentials in per-capita local government expenditure is merely an outcome of market disequilibrium. In the long run, they assert, the supply of housing will adapt to bring the market in to equilibrium, removing any short-run under or oversupply of local public services.

This theoretical critique is however far from realised in practice, as the supply of housing in particular geographies remains restricted even in the long-run. Hamilton (1983) suggests that residents fiercely resist local housing developments, aware that increased supply reduces the price premium that their own properties can command for favourable neighbourhood characteristics. More recently, Dixon and Adams' (2008) assessment of the shortage of brown-field land combined with widespread resistance to green-belt developments highlights further still the practical constraints restricting the supply of new housing. If then, the supply of housing is restricted even to some extent in the long-run, we find ourselves away from the perfectly competitive equilibrium hypothesised by Edel and Sclar. In such a case, the Tiebout model remains an important instrument for the assessment of housing markets.

In a less fundamental, but no less significant critique, Epple and Zelenitz (1981) and Wales and Wiens (1974) argue that Oates's results are spurious in nature, as his model is unable to properly control for house and neighbourhood characteristics. Whilst this assessment decreases the validity of Oates's reported coefficients, it does not undermine his underlying methodology. Their critique implies that the model is potentially of use if sufficient controls are added to account for other house price determinants. This goal, to develop a robust hedonic model of house prices

with both high explanatory power and ease of application, has received much attention in the ensuing literature.

An influential paper came in 1974 when Sherwin Rosen outlined formally a detailed theoretical model of hedonic analysis. The paper, whilst not applying the theory to a dataset itself, remains the most widely cited article in the literature, with Rosen's model providing the theoretical base for almost all subsequent analysis. In line with Tiebout's hypothesis that house prices are a function of house characteristics and the provision of local public services, Rosen argues that overall utility is maximised by purchasing a product offering a desired mixture of component features. He explains how first-step regression analysis can be used to break down observed prices into component factors. It is this method which has been widely adopted in the literature, seeking to assess the determinants of house prices, primarily with a US focus.

## *2.2 Methodological Developments*

Early studies in the field remained primitive in nature, lacking the capacity to fully control for house and neighbourhood characteristics when assessing the impact of a chosen variable on house prices. Black (1999) argued that a lack of sufficient control for neighbourhood quality meant that estimates of the impact of local public services on house prices were often biased upwards. Pioneering a 'boundary technique,' Black compared prices of houses on opposite sides, but close to school district boundaries in Massachusetts. She argued that houses close to the boundary would have comparable neighbourhood characteristics but children would attend different schools, allowing her to isolate the impact of school quality on house prices. Black concluded that a 5 percent increase in elementary school test scores would lead to a 2.1 percent increase in district house prices, around half the value of what previous studies had estimated, highlighting that controlling for neighbourhood characteristics is vital to pursuing unbiased coefficients.

Two limitations can be identified in Black's study. Her use of school-district-, rather than individual school test-scores as a unit of comparison ignores all variation within a school district. In addition, her only measure of the quality of local education is elementary school maths tests-scores. She pays no attention to high-school/secondary-school performance and focuses on only one subject. Her study essentially relies on the untested and unstated assumption that elementary maths tests-scores are an appropriate proxy for the quality of local education, a questionable conjecture. While the boundary technique provided an important development in the literature in emphasising the importance of neighbourhood characteristics, clearly Black's study left room for improvement.

Downes and Zabel (2002) add to the literature with their study of the impact of school performance on house prices in Chicago. Rather than adopting a boundary technique, Downes and Zabel control directly for neighbourhood characteristics by explicitly including a number of measures of neighbourhood quality in the regression. They overcome the first of Black's limitations by using data at the individual school level, rather than at the district level, increasing the explanatory power of the model. This aligns with research by Fletcher *et al.* (2000), who assert that better results are obtained by modelling at a disaggregated level. Similarly, to Black, though, Downes and Zabel continue to use elementary school maths tests scores as a measure of school performance, failing to explain why they are considered to be best proxy for school performance in their study. Their findings state that a one percent increase in test scores leads to a one percent increase in house prices.

Despite being far from faultless, their study should be commended for its efforts to control for neighbourhood characteristics directly, using a complex dataset at a far more disaggregated level that was seen in previous studies. The methodology perused by Downes and Zabel (2002) can be seen as a small but important innovation in the literature. The use of explicit controls for neighbourhood characteristics has since gained prominence; with subsequent studies opting for Downes and Zabel's adaptation of Oates's (1969) theory rather than Black's (1999). This dissertation will introduce neighbourhood variables in to the regression equation to control for neighbourhood characteristics, thereby adopting Downes and Zabel's preferred methodology.

### *2.3 UK Application*

The above methodology, whilst popular in the US, has rarely been applied to a UK context, with only a handful of papers - reviewed below - applying hedonic price analysis to assess the impact of school performance on UK house prices. Given that its application to the UK context is particularly scarce, wide gaps in terms of geographical location studied and flaws in methodology leave cavities in the literature, which the present research contributes towards filling. This sub-section will assess two prominent UK studies, finding significant scope for methodological improvement.

Gibbons and Machin (2003) assess the impact of English primary school quality on house prices, finding that a one percentage point increase in the proportion of pupils meeting KS2 government targets leads to a 0.67 percent increase in property prices. They argue that "mean neighbourhood property prices and mean neighbourhood school performance will provide just as much information as data based on individual schools and catchment areas" (pp.201-202), classing a 'neighbourhood' as a postcode sector. This dissertation argues that this technique is a step backwards in the literature, as aggregating data at the postcode sector omits vital information

from the dataset. This contrasts with efforts in the US literature which typically strive to utilise data at the most disaggregate level available, to better control for local area characteristics.

A more comprehensive approach was taken by Rosenthal (2003), who utilises data from both Ofsted school reports and GCSE results to measure for school quality. In addition, Rosenthal controls for neighbourhood characteristics using ACORN neighbourhood classifications, which measures neighbourhood quality at a more disaggregated level than the postcode sector (CAI, 2015), improving on Gibbons and Machin's approach. Rosenthal concludes that a 10 percentage points increase in the proportion of pupils attaining 5 A\*-C's at GCSE level leads to a 0.5 percent increase in house prices. This coefficient is particularly small compared to US estimates, although Rosenthal claims this is a result of her superior methodology. A key shortcoming of Rosenthal's study however is one which is prevalent in almost all studies conducted to date: houses are assigned to schools using only 'straight line distance to the nearest school' methodology. Her admission that only 87 percent of students actually attend their nearest secondary school shows that error has been introduced into her dataset. This dissertation overcomes this source of error by using a distinct catchment area approach, detailed in full in Section 3, after a discussion of a hedonic house price model in the geography of interest: Sheffield.

#### *2.4 Hedonic Analysis in Sheffield*

Given that much of the literature continues to find that geographical location plays a central role in determining house prices, it is important to consider the geography in which this study is located. A search of academic databases finds that very few studies have applied hedonic house price analysis in the city of Sheffield. The most cited of the few studies available is the one by Henneberry (1998). Although the study does not model the impact of school performance specifically, it needs to be discussed as its findings might highlight specific locational factors, which might have to be considered in this dissertation.

Henneberry adopts a typical hedonic house price equation to assess the impact of the Sheffield Supertram light rail system on local house prices. He finds that despite some short run effects on house prices during construction, the presence of the Supertram had no statistically significant impact on house prices two years after it had become operational. Henneberry's study can be considered a poor model of house price determinants in Sheffield and can be critiqued from two angles. First, the study can be criticised for using property asking prices rather than transaction prices. This necessarily introduces error to the model, and even Henneberry concedes that discrepancies between the two values are often present. Secondly, Henneberry uses area dummies to control for neighbourhood characteristics. The dummies provide no indication as to which

specific features of the neighbourhood are actually driving house prices and are assigned by using subjective local 'knowledge' about which areas constitute particular neighbourhoods, limiting the applicability of his model to geographies other than Sheffield. For these reasons, Henneberry's study is of little use for this dissertation, other than to highlight that the Supertram need not be included as an explanatory variable.

### 3. Model Specification and Data Sources

This section will outline the data used in this study, together with a presentation of the hedonic price equation. The key issues raised in the literature review will be addressed, including a discussion of the rationale behind choosing and omitting certain variables. Summary statistics and preliminary data analysis will also be presented.

#### 3.1 The Model

As discussed, the transaction price of a house (TP) is a function of both its physical characteristics (P) and neighbourhood characteristics (N). So that:

$$TP = f(P, N) \quad (1)$$

'Neighbourhood characteristics' are then broken down to isolate the impact of local school quality. The hypothesis is that transaction prices are positively related to the quality of local education (LE), so that:

$$TP = f(P, N, LE) \quad (2)$$

This hypothesis is tested using OLS regression analysis. The equation takes the following form:

$$\ln(TP) = \beta_0 + \beta_1(P) + \beta_2(N) + \beta_3(LE) + \epsilon \quad (3)$$

Where  $\ln(TP)$  is the natural-log of the transaction price,  $\beta_0$  is a constant, P is a vector of physical property variables, N is a measure of neighbourhood characteristics, LE is a measure of the quality of local education and  $\epsilon$  is the error term.

The model therefore takes a log-linear form, as is typical in the literature. Appendix A shows that the distribution of the natural log of sold price approximates to a normal distribution, a desirable condition for OLS regression analysis (Stock and Watson, 2012). The terms sold price and transaction price will be used interchangeably.

The data is treated as cross-sectional and data points are assumed to be independently and identically distributed, meaning that OLS analysis minimises the sum of the squared residuals. Heteroscedasticity,



multicollinearity, and the presence of outliers, all of which could undermine the performance of OLS as the chosen regression technique, are tested for and discussed throughout Section 3 and 4.

### 3.2 Discussion of Variables

#### 3.2.1 Dependent Variable

The primary source for data collection was [www.rightmove.co.uk](http://www.rightmove.co.uk). Rightmove features adverts from local and national estate agents to advertise properties for sale. They offer the largest selection of new build and resale homes in the UK, with around 90 percent of all sold property advertised on their site. In addition, when properties do sell, Rightmove combines data from the Land Registry with information contained in the initial property advert to provide a detailed database of sold properties across the UK. The ‘Sold Prices’ section of their website, from which data for this study was gathered, includes the following information: Sales Date, Transaction Price, Property Type, Full Address and Postcode, Land Ownership Type, Key Features, Floorplan and Images (Rightmove, 2015). This allows for a rich dataset to be collected, providing much more detailed property information than if data were collected from the Land Registry alone.

The sample consists of all of those properties with sufficient data available that were registered as sold between the 23<sup>rd</sup> October and 25<sup>th</sup> November 2015 inclusive. Properties where only insufficient data was obtainable (e.g., where a floorplan was missing), were excluded from the sample, resulting in a final sample size of 251 properties. Using the Land Registry's (2016) analysis tool, it is seen that house prices in Sheffield rose by 0.75% from October 2015 to November 2015. Given this, and the fact that the sample covers only 33 days, we will assume that house price inflation within the period is negligible, and the sample will be treated as a cross-section.

Summary statistics for sold price and the natural log of sold price are presented in Table 1.

**Table 1:** Distribution of Sold House Prices

Variable	Mean	Std. Dev.	Min	Max
<b>Sold Price</b>	169928.1	105263.3	30000	725000
<b>Ln(Sold Price)</b>	11.892	0.537	10.31	13.494

Source(s): Rightmove (2015)

#### 3.2.2 School Performance Variable

Oates’ (1969) influential study, as already discussed, used school inputs (per-pupil expenditure) as a measure of the provision of local education. Subsequent research has however since

questioned this approach. A widely cited paper by Hanushek (1986) examines the relationship between schooling inputs and outputs, finding little correlation exists between the two variables; per-pupil expenditure was found to be unrelated to standardised test scores for example. Despite extensive research following Hanushek's findings, Taylor and Nguyen (2006) note how there remains no consensus in the literature as to the extent at which educational outcomes are determined by inputs. Given that contemporary literature typically favours school output measures, as inputs are (at best) only loosely related to actual school performance, this dissertation will use an output measure as the variable of choice for the quality of education.

The question then arises as to which output measure is the most appropriate indicator of school quality. In addressing this question it is important to consider both the existing literature and the ultimate function of the variable in the context of this research.

Arguments for using GCSE (General Certificate of Secondary Education) test scores typically centre around the fact that studies continue to find correlation between attainment at GCSE level and positive labour market outcomes, including better jobs, and a pay premium for high achievers (Deyra, 2015). Furthermore, McIntosh (2006) notes how GCSE qualifications are the primary gateway for further study both at Advanced-Level and ultimately Higher Education. These arguments support the use of GCSE test scores, as parents are likely to be aware of the positive impact that attaining good GCSEs can have on pupils' future prospects.

Despite the popular use of GCSE results as a measure of educational performance, many education economists, including Meyer (1997) and Wilson (2004), argue that 'Value-Added' is a superior measure of school performance. Whilst this view has gained prominence in the literature when assessing changes in school quality over time, this dissertation argues that the variable of interest for this study is not necessarily *actual* school quality, but *perceived* school quality. Capitalisation of school performance within local property prices will occur in catchment areas of what parents perceive to be good schools. The success of this research relies on selecting a variable to proxy for perceived school performance, to which our attention now focuses.

Gibbons and Silva (2011) find that parents' perceptions of school quality are strongly related to standardised test scores, and only moderately correlated with their child's actual enjoyment and happiness at school. Similarly, a qualitative study by Holme (2002) found that parents were less concerned with newer, holistic measures of performance, and more concerned with historical reputations and schools which produced 'high-achievers', both of which are typically associated with good performance on standardised testing.

Given these findings, this study will use the percentage of pupils achieving 5A\*-C grades at GCSE level as a measure of school performance. The variable is summarised in Table 2, where it can be seen that the best performing school (Tapton) had 82 percent of students meeting the criteria, whilst for the worst performing school (Chaucer) the figure was only 26 percent.

**Table 2:** Distribution of School Performance Variable

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
<b>School Performance</b>	54.586	13.486	26	82

*Source(s):* Department for Education (2015)

Houses have been assigned by the author to schools using specific catchment areas, by using the locator tool on Sheffield City Council's website (Sheffield City Council, 2015). This (highly labour-intensive) approach is distinctly different from the 'nearest-school' approach common in the literature. Appendix B and C show how secondary school catchment areas are markedly non-uniform in nature. In this way, the present study is distinct from the existing literature. In capturing data on actual catchment areas, the significant errors introduced by the 'nearest school' methodology are avoided.

Sheffield City Council run a catchment area-based admissions policy, full details of which are published online (Sheffield City Council, 2016). To summarise, for oversubscribed schools, the catchment area is the second most important criteria in allocating places, after first priority is given to students in local authority care. All academies which were formerly under local authority control have adopted the same admissions policy as the Council (Sheffield City Council, 2016). The study does not include schools without defined catchment areas, such as faith schools, or special educational needs schools. None of the schools in the study select students based on academic achievement.

### *3.2.3 Property Control Variable*

Controlling for house characteristics is central to the success of the model. Data was first collected on the property type, with each observation assigned to the category 'flat,' 'terraced,' 'semi-detached' or 'detached'. These variables entered the regression as dummy variables, with the variable 'flat' omitted as the base. Further dummy variables included were 'garage,' 'conservatory,' 'good-condition' and 'bad-condition'. The base category for condition of the property was 'average-condition'. A subjective judgement based on images of the property was

used to assign properties to a 'good' or 'bad' classification, together with estate agents' comments in advertisements such as "requires modernisation" or "beautifully presented". In cases where any doubt remained, houses were assigned to the 'average' category. The number of bedrooms and the number of bathrooms are also included; these are the only continuous property control variables.

**Table 3:** Property Control Variables - Summary Statistics

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
<b>Flat</b>	0.056	0.230	0	1
<b>Terraced</b>	0.382	0.487	0	1
<b>Semi</b>	0.430	0.496	0	1
<b>Detached</b>	0.131	0.339	0	1
<b>Garage</b>	0.375	0.485	0	1
<b>Conservatory</b>	0.155	0.363	0	1
<b>Good Condition</b>	0.175	0.381	0	1
<b>Bad Condition</b>	0.191	0.394	0	1
<b>No of Bedrooms</b>	2.960	0.784	1	6
<b>No of Bathrooms</b>	1.139	0.369	1	3

*Source(s):* Rightmove (2015)

From Table 3 it can be seen that the largest subset of property type was semi-detached houses. The majority of houses did not have a garage, and most houses did not have a conservatory. Less than six percent of properties were flats, whilst the average number of bedrooms was approximately three. Positive coefficients are expected on all property control variables, other than 'bad-condition', for which a negative coefficient is expected.

Data was also collected on the land ownership of the property. The impact of a property being freehold or leasehold typically had a significant effect in reported regressions. Sheffield can however be seen as an anomaly regarding land ownership; many properties are technically leaseholds but the ground rate is a nominal fee to a historical landlord (Mortimore, 1969). Summary statistics are presented in Table 4. The difference between the two means was not statistically significant (P-value = 0.8). Land ownership was therefore not included in any regression equation.

**Table 4:** Land Ownership - Summary Statistics

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>No of Observations</b>
<b>Freehold</b>	171596.8	119887.7	123
<b>Leasehold</b>	168324.6	89438.45	128

Source(s): Rightmove (2015)

### 3.2.4 Neighbourhood Control Variable

As outlined in Section 2, controlling for neighbourhood characteristics is a fundamental requirement for a successful model. Studies often link the prevalence of crime to the general quality of a neighbourhood, with research finding that crime and neighbourhood affluence are negatively related (i.e. poorer areas have higher crime rates) (Gyimah-Brempong, 2006; Eriksson, 2016). Data was therefore collected on local crime statistics to act as a proxy for neighbourhood quality. The measure recorded was the number of crimes reported in November 2015 within a one-mile radius of a property's address and was gathered from the Police UK website (College of Policing Limited, 2015). The variable 'local crime' had a mean of 305 and a range from 18 to 1452. As detailed in Section 4.1, regression estimation tested alternative indicators as proxies for 'neighbourhood quality' and ultimately 'local crime' did not feature in the final preferred equation.

An alternate measure of neighbourhood quality was also collected. The English Indices of Multiple Deprivation (IMD) is a dataset compiled by the Department for Communities and Local Government (2015). The IMD ranks English neighbourhoods from the most to the least deprived, where small area number 1 is the most deprived, and small area number 32844 is the least deprived. Small areas are defined as Lower Layer Super Output Areas (LSOA), the smallest area at which government statistics are produced. The IMD Research Report (Smith *et al.*, 2015a) notes how each LSOA typically has around 1,500 residents, significantly fewer than at the postcode district for example. Given that the IMD takes into account a number of measures of deprivation, it is hypothesised that it will be a useful dataset to control for the quality of an area at a highly disaggregated level.

The overall IMD ranking is made up of seven weighted sub-domains of deprivation (percentage weights): Income (22.5), Employment (22.5), Education Training and Skills (13.5), Health and Disability (13.5), Crime (9.3), Living Environment (9.3) and Barriers to Housing and Services (9.3).

A measure of the unedited IMD statistics (which groups areas by decile) was collected and named 'IMD decile'. This explanatory variable did not feature in the final preferred specification as the chosen proxy for neighbourhood quality due to multicollinearity issues discussed below.

The Education Training and Skills component of the IMD includes a range of indicators to measure the skills and attainment of both children and adults, including data on GCSE performance (Smith *et al.*, 2015b). Given that this component is correlated with the variable 'school performance' its inclusion could lead to multicollinearity and a large standard error for the variable of interest (Stock and Watson, 2012). It is for this reason that the raw IMD data has been manipulated to exclude the Education Training and Skills component whilst retaining the relative weightings of the remaining elements. The variable is named 'adjusted IMD'. A positive coefficient is expected, so that less deprived neighbourhoods are correlated with a higher transaction price. Appendix D shows a scatter plot of Ln(Sold Price) against 'adjusted IMD,' showing a linear relationship between the two variables.

### *3.3 Detecting Multicollinearity*

Appendix E presents a correlation matrix between all variables introduced in Section 3.2. Only one measure of neighbourhood quality will be used in each regression equation, due to the overlap and correlation between 'local crime,' 'IMD decile' and 'adjusted IMD'. Other than between the different proxies for neighbourhood quality, it can be seen that none of the explanatory variables are highly correlated with each other, suggesting that it is unlikely that the model will suffer from the issues associated with multicollinearity. The highest correlation between two property control variables was 0.57, between the variables 'no of bathrooms' and 'no of bedrooms'. When analysing reported coefficients care will be taken to check the significance of these two variables, as multicollinearity would increase the standard deviation of the coefficients of the correlated variables.

## 4. Estimation and Results

### 4.1 Estimation

Stock and Watson (2012) note the importance of beginning estimation with a basic regression model with a limited number of variables, and to then build a complex model sequentially. The benefit of this is to see the impact of an additional variable on the model and understand how and why it has affected the results. The ensuing discussion will explain how the final preferred equation has been chosen, why certain variables have been included/excluded and how the issue of omitted variable bias has been mitigated. The entirety of Section 4 relates to the data presented in Table 5.

Before exploring the impact of school performance on house prices, the first estimations explore the basic drivers of house prices according to property and location characteristics. Analysis begins by regressing  $\ln(\text{Sold Price})$  against a handful of basic property features, including the property type, number of bedrooms, and whether or not the property featured a garage or conservatory. The low adjusted R-squared shows that the model does a poor job at explaining the variance in the data. Coefficients on 'terraced,' 'semi-detached,' 'garage' and 'conservatory' were not significant, although these variables will remain in the model for the time being, as intuition suggests the variables should be important household characteristics.

Regression 2 includes the variable 'no of bathrooms'. This reduces the coefficient on the variable 'no of bedrooms,' showing that in specification 1 the coefficient on the number of bedrooms was biased upwards by acting as a proxy for the number of bathrooms. Despite some correlation between the two variables (identified in Section 3.2), both remain significant at the five percent level, indicating that multicollinearity is not a critical issue in this instance. Although specification 2 only slightly improves the adjusted R-squared value, the introduction of a variable to account for the number of bathrooms is a welcome addition as it reduces the omitted variable bias in the model.

Regression 3 completes the introduction of property control variables. The introduction of 'good-condition' and 'bad-condition' controls increase the fit of the model by ten percent, though the coefficient on the number of bathrooms is no longer significant. The presence of a garage does for the first time become significant, although only at the 10 percent level. Comparing regressions 1 through to 3 we can see that the addition of variables to more thoroughly account for property characteristics has increased the fit of the model and reduced the amount of omitted variable bias. Non-significant variables will continue to be included as they may become significant once neighbourhood characteristics have been properly accounted for.

Regressions 4-6 examine different controls for neighbourhood quality. Specification 4 builds on the previous by introducing 'local crime' as a measure of local area quality. The variable is significant at the one percent level and increases the adjusted R-squared value above model 3. A coefficient of  $-0.3 \times 10^{-3}$  implies that an increase in crimes within a one-mile radius of the property of 100 per month corresponds to a three percent decrease in price. This suggests that crime statistics may be a valuable indicator of neighbourhood quality.

Regression 5 drops the 'local crime' variable, instead using variable 'IMD decile'. This increases the fit of the model by a considerable amount. The coefficients on the variables 'no of bedrooms,' 'detached' and 'good-condition' reduce compared to model 3. These variables are thought to have been proxying slightly for the characteristics of the neighbourhood in the earlier model (i.e., nicer, bigger houses with more bedrooms tend to be in nicer areas). Given the superior increase in adjusted R-squared that model 5 offers over model 3, compared to model 4 over model 3, we can confirm that the 'IMD decile' is a better measure of neighbourhood characteristics than crime rates in that it has more explanatory power.

Given that regression 5 highlights the potential use of the IMD dataset, model 6 drops 'IMD Decile' and introduces the variable 'adjusted IMD' (described in Section 3.2.4). The 'adjusted IMD' variable retains more explanatory power by not grouping results in to deciles and represents better statistical practice as it has reduced correlation with 'school performance'. Its introduction to the model increases slightly the adjusted R-squared value, and coefficients on other control variables changed only negligibly when compared with those of model 5. The property and neighbourhood control variables in specification 6 offer an improvement over earlier models and will be taken forward to assess the impact of school performance.

This then provides the preferred model of Sheffield house prices based on property and area characteristics. Next, analysis turns to the impact of the key variable of interest: 'school performance'. Does this provide an independent effect on house prices over and above other characteristics? The first results are shown in regression 7. The 'school' variable is statistically significant at the one percent level and increases the adjusted R-squared value. Its addition causes the coefficient on 'terrace' to double, though it falls short of being significant at the 10% level (t-statistic=1.39). Meanwhile the coefficient on 'semi-detached' increases and becomes statistically significant at the 5% level, while the coefficient on the variable 'detached' increases from 0.44 to 0.54. This indicates that in previous models both 'semi-detached' and 'detached' were biased downwards by not accounting for the impact of school performance. In addition, the coefficient on the variable 'adjusted IMD' is reduced by 19 percent compared to model 6, indicating that it was



previously biased upwards by partly acting as a proxy for school performance. The increase in the adjusted R-squared value indicates that the introduction of 'school performance' has improved the explanatory power of the model and is thus a justified introduction.

Regression 8 tests the hypothesis that the relationship between  $\ln(\text{soldprice})$  and school performance is non-linear. A squared term for school performance variable is added. Results state that the coefficients of 'school performance' and ('school performance')<sup>2</sup> are both non-significant, with the coefficient of 'school performance' actually being negative. The value of the adjusted R-squared is increased only by a trivial amount. These results indicate that the addition of the squared term has not improved the model and thus regression 8 is not an improvement over specification 7.

The preferred equation is regression specification 7. It is able to control for property and neighbourhood characteristics better than any other tested model, allowing for assessment the impact of school performance on house prices. Whilst the variables 'terraced,' 'garage' and 'conservatory' continue to be not significant at the 10 percent level, they remain in the equation as both intuition and previous literature suggest they should not be excluded from a final model. An interpretation of the results will be provided in Section 4.2, before post-estimation diagnostics are performed in Section 4.3.

**Table 5:** Regression Results of the Natural Log of Property Prices on School Performance Taking Account of Property and Neighbourhood Characteristics;  
Dependent Variable: Ln(Sold Price)

Independent Variable:	1	2	3	4	5	6	7	8
<b>Terraced</b>	-0.0532	-0.0153	0.0006	-0.0389	0.05	0.0616	0.1262	0.1272
<b>Semi</b>	0.0857	0.1312	0.1107	0.0489	0.1405	0.1474	0.224**	0.2224**
<b>Detached</b>	0.4814***	0.4852***	0.5005**	0.4121***	0.4266***	0.4426***	0.5414***	0.5354***
<i>Flat (reference category)</i>								
<b>No of Bedrooms</b>	0.3042***	0.2557***	0.2663***	0.285***	0.2185***	0.2160***	0.1890***	0.1868***
<b>No of Bathrooms</b>		0.1886**	0.103	0.0966	0.1425**	0.1375**	0.1272**	0.1416**
<b>Garage</b>	0.0884	0.0808	0.1060*	0.0819	0.0517	0.0367	0.0227	0.0172
<b>Conservatory</b>	0.087	0.07	0.0757	0.0535	0.0523	0.0446	0.065	0.0666
<b>Good Condition</b>			0.1745***	0.1619**	0.0940*	0.0948**	0.1151**	0.1141**
<b>Bad Condition</b>			-0.2151***	-0.2073***	-0.1462***	-0.1441***	-0.1368***	-0.1324***
<b>Average Condition</b> <i>(reference category)</i>								
<b>School Performance (*10<sup>-3</sup>)</b>							7.8515***	-4.2711
<b>School Performance<sup>2</sup></b> <i>(*10<sup>-3</sup>)</i>								0.107
<b>IMD Decile</b>					0.0765***			
<b>Adjusted IMD (*10<sup>-3</sup>)</b>						0.0302***	0.0245***	0.0250***
<b>Local Crime (*10<sup>-3</sup>)</b>				-0.301***				
<b>Constant/Intercept</b>	10.865***	10.7647***	10.8317***	10.9415***	10.4974***	10.4441***	10.1157***	10.4326***
<b>Summary Statistics:</b>								
<b>R-squared</b>	0.4556	0.4658	0.5115	0.5268	0.6788	0.6815	0.7112	0.7132
<b>Adj R-squared</b>	0.4422	0.4504	0.4933	0.5071	0.6654	0.6682	0.6979	0.6988
<b>N</b>	251	251	251	251	251	251	251	251

Note: The coefficient is significant at the \*10% level, \*\*5% level or \*\*\*1% using a two sided test.

Source(s): Author's research, using STATA statistical software. See Section 3 for individual variable sources.

#### *4.2 Interpretation of Results*

Analysing results from the favoured specification it is seen that a one percentage point increase in the proportion of pupils achieving 5A\*-Cs at GCSE level will increase house prices in the corresponding school catchment area by 0.785 percent. The variable is statistically significant at the one percent level (t-statistic = 4.94), indicating that the likelihood of the coefficient actually being equal to zero is well below conventional thresholds. These results show that school quality is an important determinant of property prices. Findings from the literature review showed that reported elasticities in prior research were: 0.42 (Black, 1999), 0.67 (Gibbons and Machin, 2003), 0.05 (Rosenthal, 2003) and 1.00 (Downes and Zabel, 2002). The 0.785 elasticity reported in this research therefore falls well within the expected range typically found in the literature.

Consider the hypothetical situation of two identical houses in identical areas differing only by the school catchment area in which they fall. If house A finds itself in a catchment area where the proportion of pupils achieving 5A\*-Cs is 45 percent, and house B finds itself in a catchment area where the corresponding figure is 55 percent, then house B is expected to be worth 7.85 percent more than house A. At the mean property price for Sheffield this would equate to around £13,340. It is also typical to consider the case of a one standard deviation increase or decrease in the variable of interest. The results from this study find that a one standard deviation increase in school performance from the mean (representing an increase from 54.6 to 68 percent of students meeting the grade) would equate to a 10.59 percent increase in value for properties within the defined catchment area. At the mean property price for Sheffield this is equivalent to £17,995.

Despite not being the direct focus of investigation, the model allows for independent analysis of the effects of property features on transaction prices. Results find that property type can have a notable impact on prices, with a detached house able to command a 26 percent price premium over a comparable semi-detached property. In contrast however, the difference between a comparable flat or terraced property was not statistically significant. The impact of an additional bedroom or bathroom is marked, with a one unit increase affecting property values by 19 and 13 percent respectively. Surprisingly, the study found that even once rigorous controls for neighbourhood characteristics were included, the effect of the presence of a garage or conservatory was not significantly different from zero. Less surprising however was the hedonic price attributed to the condition of the property: houses in good condition commanded around a 10 percent premium, whilst properties in poor condition saw typical prices 14 percent lower than the 'average' category.

With regards to neighbourhood characteristics, the 'adjusted IMD' variable controlled well for area quality, increasing the fit of the model considerably more than controlling for property features alone. Estimates show that the coefficient on the 'adjusted IMD' variable was biased upwards in specification 6 by the non-inclusion of a measure of school performance. Regression 7 therefore contributes to the field by highlighting the importance of school catchment areas as an independent factor distinct from simple 'neighbourhood features'. The coefficient on 'adjusted IMD' sadly has little practical interpretation. The simplistic understanding is that more deprived areas are associated with lower house prices - a predictable outcome. It is beyond the scope of this study to further break down the neighbourhood effects beyond what has been done so in isolating the impact of school performance.

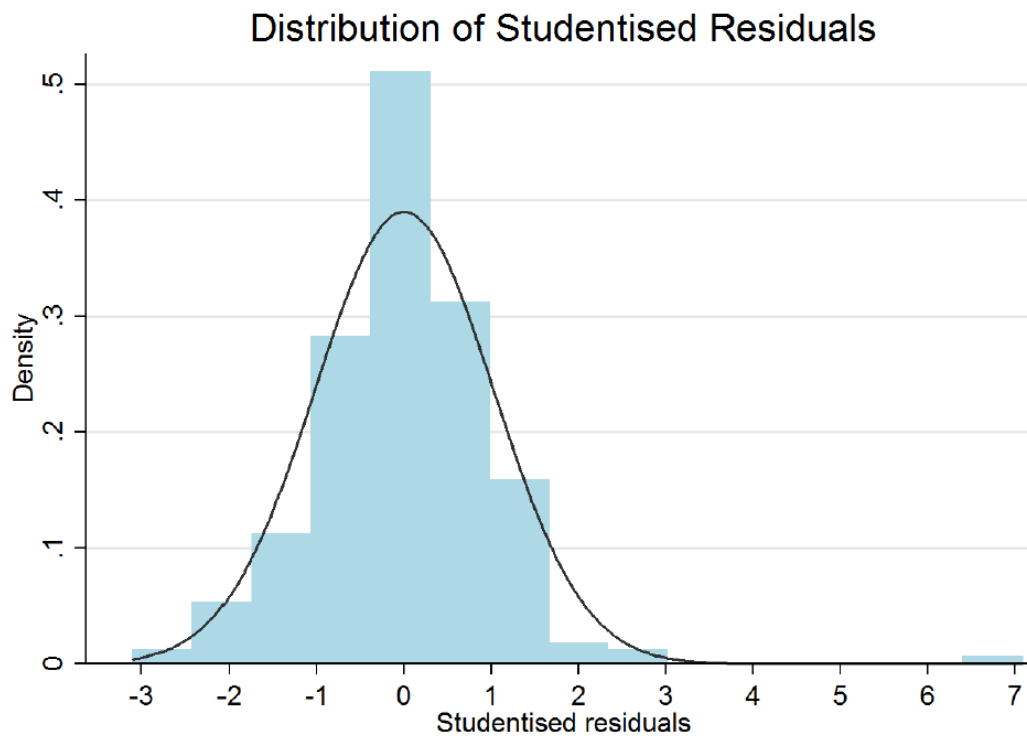
#### *4.3 Post-estimation Diagnostics*

This section will find that the preferred model is robust to a number of sensitivity checks, and thus the stated coefficients and their implications can be reported with a high degree of confidence.

Post-estimation analysis begins by testing for homoscedastic error terms using a Breush-Pagan/Cook-Weisberg test for heteroscedasticity. The null hypothesis, that errors are homoscedastic, is not rejected by the test, so we are able to conclude that the variance of the error terms does not change in a linear fashion with  $\ln(\text{Sold Price})$  (chi-squared value = 2.35; critical value = 14.07).

The functional form of the model is then tested using a Ramsey RESET test to ascertain if the relationship between the dependent variable and the independent variable is indeed linear. The alternative hypothesis is that the correct specification is non-linear. Running the test in STATA shows that the F-test-statistic is 0.47 (P-value = 0.7). This results in a rejection of the alternative hypothesis that the model is wrongly specified, confirming that a linear model is the correct functional form.

Figure 1:



Source(s): Author's calculations using STATA statistical software.

In order to determine if outliers are driving the results of the model, we first examine the residuals, a technique well applied in the literature and summarised well by Anscombe and Tukey (1963). Outliers represent points that are far away from the model's prediction and will have a residual large in absolute magnitude. In order to compare how unusual individual outliers are, studentised residuals are calculated, the value of which is a residual divided by its standard deviation. Figure 1 plots the distribution of the studentised errors, with a normal distribution superimposed. It is generally accepted that a studentised residual value greater than 3 in absolute magnitude is considered an outlier. It is seen from Figure 1 that the distribution approximates to normality, and there appears to be no widespread presence of outliers. There is however one residual which reports a studentised value of 7.08, representing an outlier.

Checks confirm that the outlier is not the result of a data entry error, nor is it from a different population, so choose not to exclude this unusual data-point entirely. Instead, the favoured specification is run through an Outlier Robust Regression (ORR). Regressions 1 to 7 use OLS methodology which applies equal weight to each observation in the dataset. ORR on the other hand, applies heavier weighting to observations closer to their predicted value, and lighter weighting to observations with large residuals. This results in extreme values having less impact on the results of the model and can be considered superior to OLS in cases where outliers are present

(McKean, 2004; Verardi and Croux, 2009). The results of regression 7 using the ORR approach are presented in full in Appendix F. To summarise, the coefficients on all control variables change only negligibly, with all variables remaining significant at the previous levels of confidence. The coefficient on the variable of interest changes marginally from 7.85 to 6.44; both are statistically significant at the 1 percent level. We can conclude from the ORR that no major errors were made in the original OLS specification 7, and the presence of outliers has only a small effect on the reported coefficients.

## **5. Conclusion**

An extensive review of the existing literature highlighted the importance of controlling for neighbourhood characteristics, and the benefits of using highly disaggregated data. Analysis of previous studies showed that despite methodological advances in US applications, little work has been done to apply best practice techniques to a UK context. In addition, the rationale for the choice of school performance measures has rarely been explained, and in many cases is questionable. The most concerning aspect of the literature however, was the apparent widespread refusal to acknowledge that a 'straight line distance to the nearest school' approach was unsatisfactory.

This dissertation questioned this consensus, by allocating houses to schools using actual catchment area boundaries. In light of Sheffield City Council's admissions policy, discussed in Section 3.2.2, this approach seems critical in avoiding large scale introduction of error in the dataset. Furthermore, the problems of controlling for neighbourhood characteristics that plagued many studies to date, were largely overcome. Data from the English Indices of Multiple Deprivation, manipulated to exclude a potential source of multicollinearity, performed well at controlling for neighbourhood factors, more so than simplistic controls such as crime statistics alone.

Findings showed that a one percentage point increase in the proportion of students attaining 5A\* - Cs at GCSE level increased house prices in the school's catchment area by 0.785 percent. The model performed well in a number of sensitivity checks, showing that reported results are robust. This 'good school price premium' seems to provide evidence to media claims that access to the best schools is restricted only to those who can afford it.

It is hoped that the regression model built in this study can be adapted and applied in further research. In the first instance, neighbourhood factors other than school performance could be further broken down to isolate the impact of the quality of the environment, or prevalence of local crime, on house prices in the city of Sheffield. The model could then be further adapted and applied to a wider UK context, to assess residents' valuations of school provision using the

methods outlined in this study. As Geographic Information System (GIS) technology develops, it is hoped that actual catchment areas can continue to be used in allocating houses to schools in a less laborious manor than was necessary in this research. Such developments would allow a bigger sample size and a wider geography than was achievable in this study, whilst retaining the key benefits gained from the catchment area approach.

## Bibliography

- Anscombe, F. J., and Tukey, J. W. 1963. The Examination and Analysis of Residuals. *Technometrics*. **5**(2), pp.141-160.
- Black, S. 1999. Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*. **114**(2), pp.577-599.
- CACI. 2015. *Acorn Technical Guide*. Kesington, London: CACI. [Online]. [Accessed: 29<sup>th</sup> February 2016]. Available at: <http://acorn.caci.co.uk/downloads/Acorn-Technical-document.pdf>
- College of Policing Limited. 2015. *Police UK Crime Map and property search tool*. [Online]. [Accessed: 1<sup>st</sup> December 2015]. Available at: <https://www.police.uk/south-yorkshire/KF4/crime/+Objva0/>
- Coughlan, S. 2016. Secondary school places under pressure. *BBC News*. 1 March 2016. [Online]. [Accessed: 14<sup>th</sup> March 2016]. Available at: <http://www.bbc.co.uk/news/education-35688548>
- Department for Communities and Local Government. 2015. *English indices of deprivation 2015: Postcode Lookup tool*. [Online]. [Accessed: 1<sup>st</sup> December 2015]. Available at: <http://imd-by-postcode.opendatacommunities.org/>
- Department for Education. 2015. *School and College Performance Tables*. [Online]. [Accessed: 1<sup>st</sup> December 2015]. Available at: <http://www.education.gov.uk/cgi-bin/schools/performance/search.pl?location=Sheffield&phase=all&searchType=location>
- Deyra, U.S. 2015. Doubly Robust Estimation of Causal Effects with Multivalued Treatments: An Application to the Returns to Schooling. *Journal of Applied Econometrics*. **30**(5), pp.736-786.
- Dixon, T., and Adams, D. 2008. Housing Supply and Brownfield Regeneration in a post-Barker World: Is There Enough Brownfield Land in England and Scotland?. *Urban Studies*. **45**(1), pp.115-139.
- Downes, T. A., and Zabel, J. E. 2002. The impact of school characteristics on house prices: Chicago 1987-1991. *Journal of Urban Economics*. **52**(1), pp.1-25.
- Edel, M., and Sclar, E. 1974. Taxes, Spending and Property Values: Supply adjustment in a Tiebout-Oates Model. *Journal of Political Economy*. **82**(5), pp.941-954.
- Epple, D., and Zelenitz, A. 2016. The Implications of Competition Among Jurisdictions: Does Tiebout Need Politics?. *Journal of Political Economy*. **89**(6), pp.1197-1217.
- Eriksson, K.H., Hjalmarsson, R., Lindquist, M.J., and Sandberg, A. 2016. The importance of family background and neighborhood effects as determinants of crime. *Journal of Population Economics*. **29**(1), pp.219-262.
- Fletcher, M., Gallimore, P., and Mangan, J. 2000. The modelling of housing submarkets. *Journal of Property Investment & Finance*. **18**(4), pp.473-487.
- Gibbons, S., and Machin, S. 2003. Valuing English primary schools. *Journal of Urban Economics*. **53**(2), pp.197-219.
- Gibbons, S., and Silva, O. 2011. School Quality, child wellbeing and parents' satisfaction. *Economics of Education Review*. **30**(2), pp.312-331.
- Gyimah-Brempong, K. 2006. Neighborhood income, alcohol availability, and crime rates. *The Review of Black Political Economy*. **33**(3), pp.21-44.



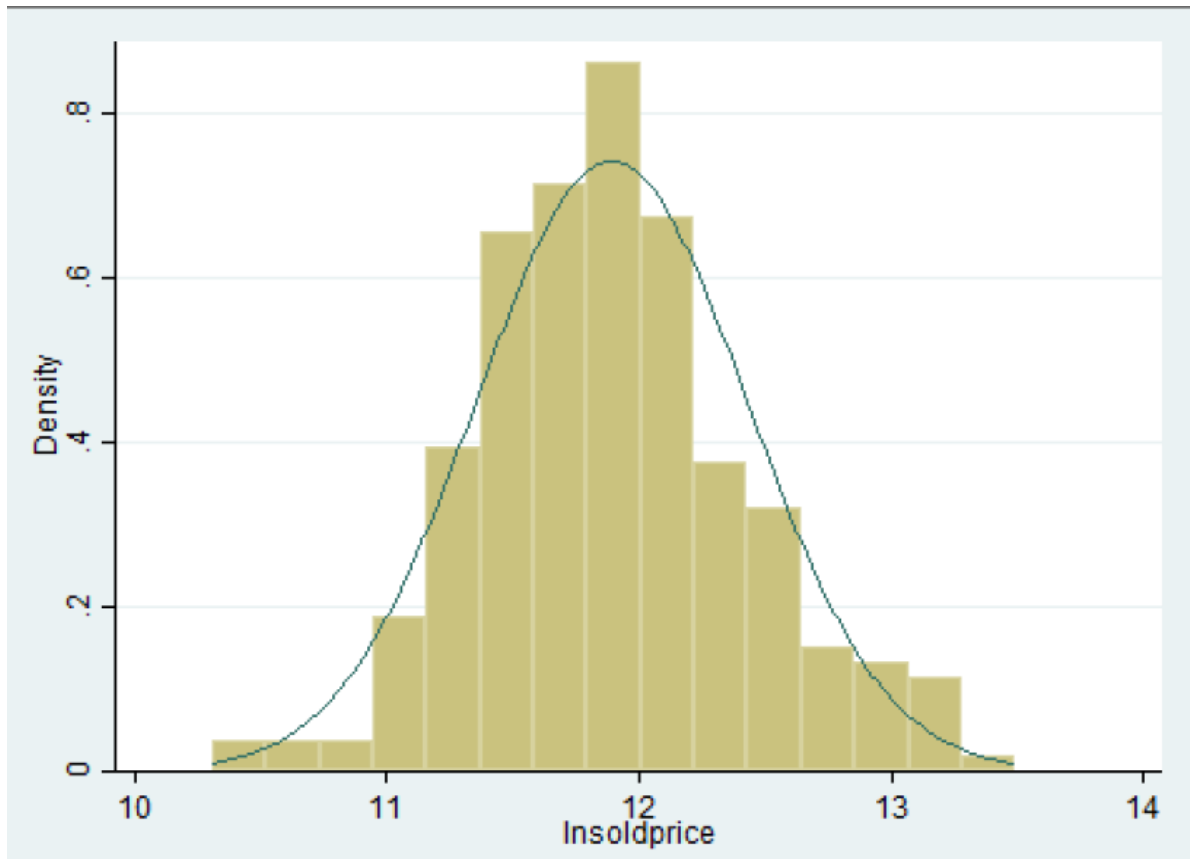
- Hanushek, E. A. 1986. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*. **24**(3), pp.1141-1177.
- Henneberry, J. 1998. Transport investment and house prices. *Journal of Property Valuation & Investment*. **16**(2), pp.144-158.
- Jones, D. 2015. £45K to get nice school. *The Sun*. 30 August 2015. [Online]. [Accessed: 14<sup>th</sup> March 2016].  
Available at: <http://www.thesun.co.uk/sol/homepage/news/6613400/Parents-will-pay-an-extra-45000-to-buy-a-home-near-a-good-school.html>
- Land Registry 2016. *Land Registry Linked Open Data*. [Online]. [Accessed: 2<sup>nd</sup> March 2016]. Available at: <http://landregistry.data.gov.uk/app/hpi/hpi>
- McIntosh, S. 2006. Further Analysis of the Returns to Academic and Vocational Qualifications. *Oxford Bulletin of Economics and Statistics*. **68**(2), pp.225-251.
- McKean, J. W. 2004. Robust Analysis of Linear Models. *Statistical Science*. **19**(4), pp.562-570.
- Meyer, R. H. 1997. Value-added indicators of school performance: A primer. *Economics of Education Review*. **16**(3), pp.283-301.
- Mortimore, M. J. 1969. Landownership and Urban Growth in Bradford and Its Environs in the West Riding Conurbation, 1850-1950. *Transactions of the Institute of British Geographers*. **46**(1), pp.105-119.
- Oates, W. E. 1969. The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis. *Journal of Political Economy*. **77**(6), pp.957-971.
- Okolosie, L. 2016. Choice in education is only for those who can afford it. *The Guardian*. 19 January 2016. [Online]. [Accessed: 14<sup>th</sup> March 2016].  
Available at: <http://www.theguardian.com/commentisfree/2016/jan/19/choice-education-afford-parent-child-failing-school-middle-class>
- ONS 2015. *Population estimates for UK mid-2014 analysis tool*. Newport: UK Office for National Statistics. [Online]. [Accessed: 1<sup>st</sup> March 2016].  
Available at: <http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesanalysistool>
- Paton, G. 2009. School admissions: the top scams. *The Telegraph*. 2 November 2009. [Online]. [Accessed: 14<sup>th</sup> March 2016].  
Available at: <http://www.telegraph.co.uk/education/6487421/School-admissions-the-top-scams.html>
- Rightmove 2015. *Rightmove: UK's number one property website for properties for sale and to rent*. [Online]. [Accessed: 1<sup>st</sup> December 2015].  
Available at: <http://www.rightmove.co.uk/>
- Rosen, S. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*. **82**(1), pp.34-55.
- Rosenthal, L. 2003. The Value of Secondary School Quality. *Oxford Bulletin of Economics and Statistics*. **65**(3), pp.329-335.
- Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., and Plunkett, E. 2015a. *English indices of deprivation 2015: Research report*. London: Department for Communities and Local Government. [Online]. [Accessed: 1<sup>st</sup> December 2015].  
Available at: <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-research-report>

- Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., and Plunkett, E. 2015b. *English indices of deprivation 2015: Technical report*. London: Department for Communities and Local Government. [Online]. [Accessed: 1<sup>st</sup> December 2015]. Available at: <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report>
- Stock, J. H., and Watson, M. H. 2012. *Introduction to Econometrics*. 3<sup>rd</sup> ed. London: Pearson.
- Sheffield City Council 2015. *Sheffield City Council Catchment Area Property Location Tool*. [Online]. [Accessed: 1<sup>st</sup> December 2015]. Available at: <https://maps.sheffield.gov.uk/LocalViewExt/Sites/SchoolCatchmentandLocation201516/>
- Sheffield City Council 2016. *Sheffield City Council School pupil admissions policy*. [Online]. [Accessed: 3<sup>rd</sup> March 2016]. Available at: <https://www.sheffield.gov.uk/education/information-for-parentscarers/pupil-admissions/admission-policy.html>
- Taylor, J., and Nguyen, N.A. 2006. An Analysis of the Value Added by Secondary Schools in England: Is the Value Added Indicator of Any Value?. *Oxford Bulletin of Economics and Statistics*. **68**(2), pp.203-224.
- Tiebout, C.M. 1956. A Pure Theory of Local Public Expenditures. *Journal of Political Economy*. **64**(5), pp.416-424.
- Verardi, V., and Croux, C. 2009. Robust regression in Stata. *The Stata Journal*. **9**(3), pp.439-453.
- Wiens, E.G. 1974. Capitalization of Residential Property Taxes: An Empirical Study. *The Review of Economics and Statistics*. **56**(3), pp.329-333.
- Wilson, D. 2004. Which Ranking? The Impact of a 'Value-Added' Measure of Secondary School Performance. *Public Money & Management*. **24**(1), pp.37-45.

## Appendices

### Appendix A

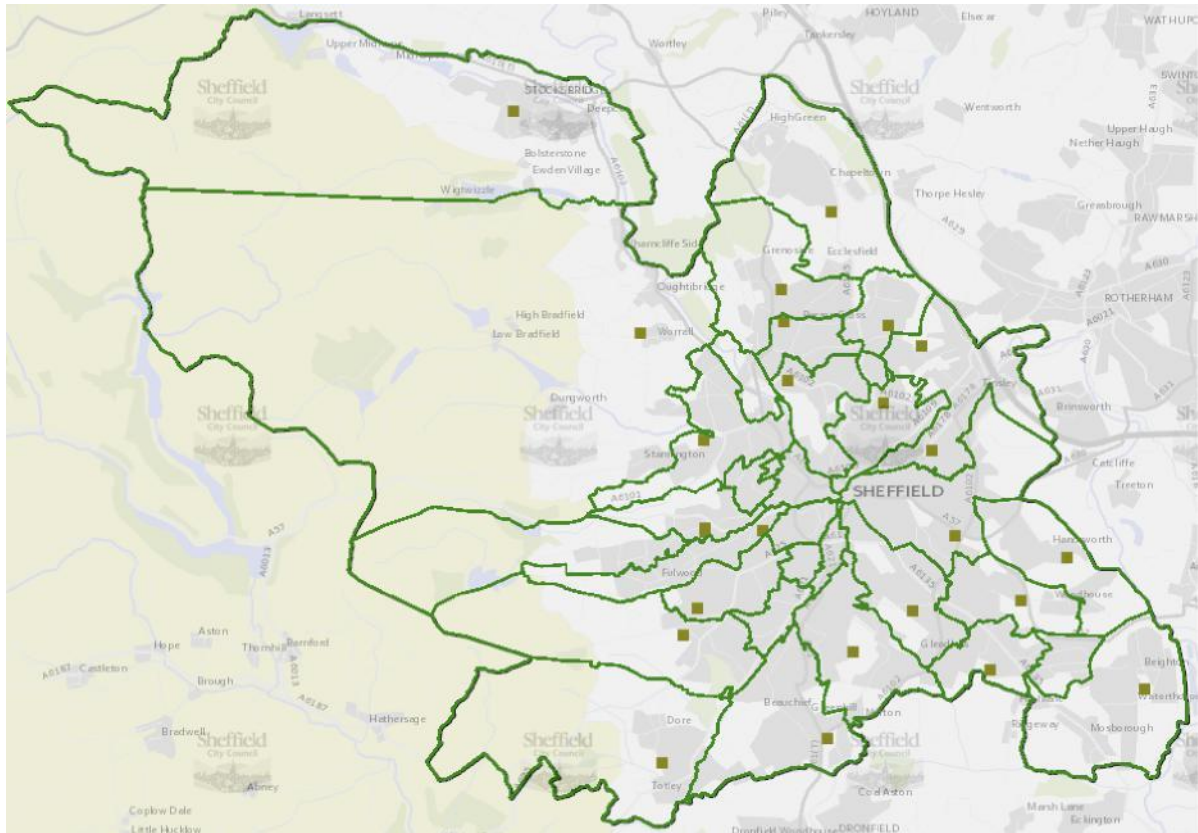
Histogram to show the distribution of the dependent variable,  $\ln(\text{Sold Price})$ , as discussed in Section 3.1



Source(s): Rightmove sold price website, 2015. Please see reference (Rightmove, 2015).

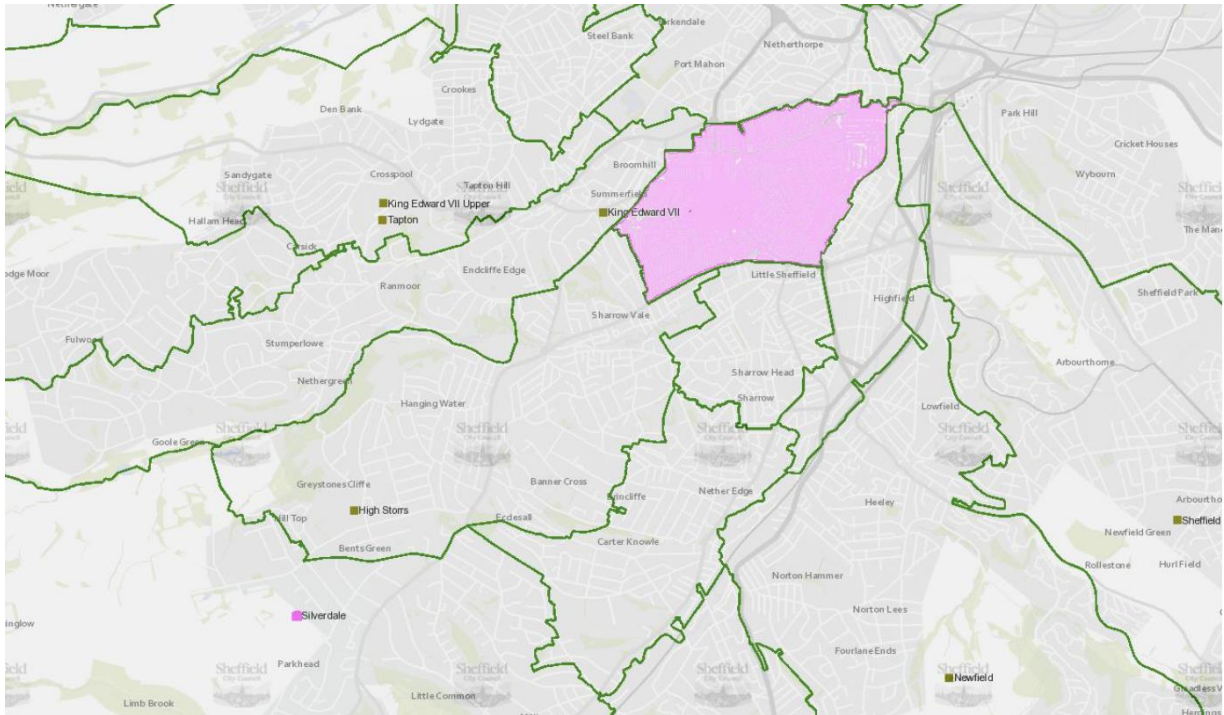
*Appendix B*

Map of secondary school catchment areas in Sheffield. Green lines represent catchment area boundaries, green squares represent secondary schools.



*Source(s):* Sheffield City Council website. See reference (Sheffield City Council, 2015).

## Appendix C

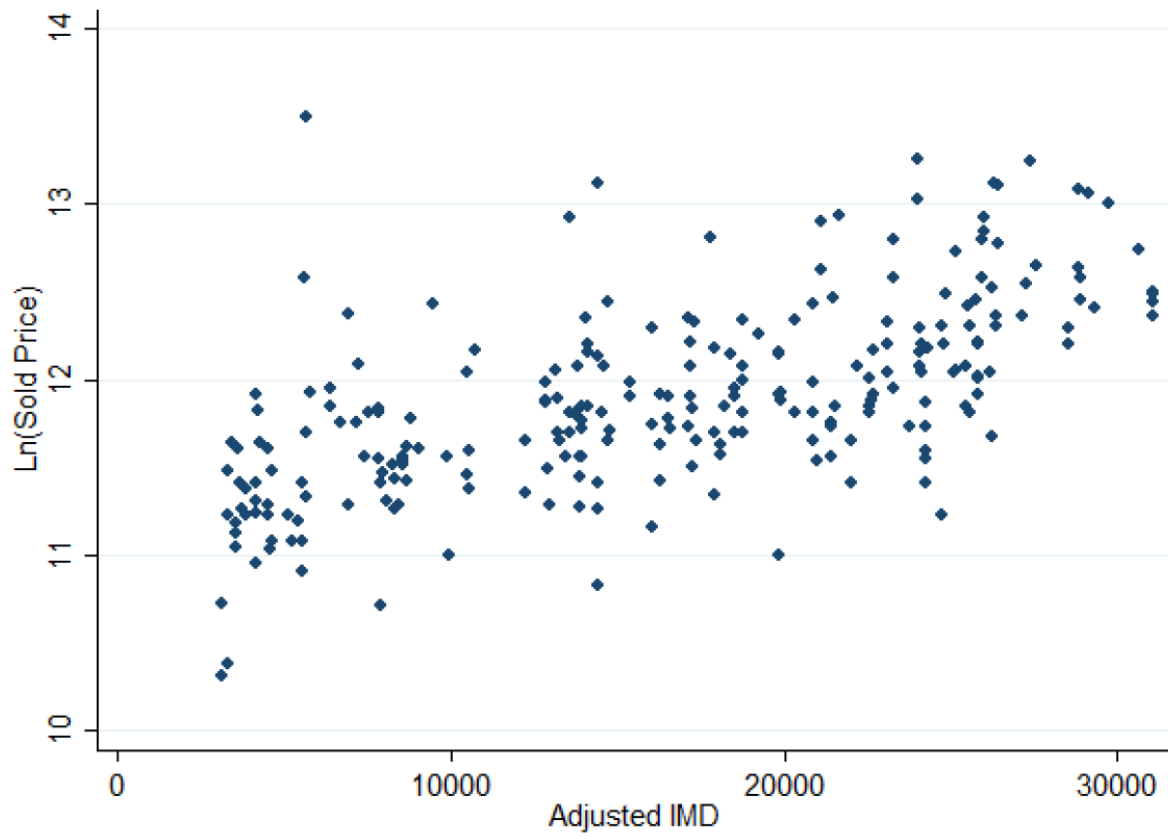


Source(s): Sheffield City Council website. See reference (Sheffield City Council, 2015).

This enlarged image of Appendix B has been highlighted to show the non-uniform nature of school catchment area boundaries. The inner-city area shaded in pink is assigned by the City Council to Silverdale School, a good school on the outskirts of the city. Pupils living in this neighbourhood live physically closer in distance to other schools, for example King Edward VIII School, but attend Silverdale School, which is also served by students living in the direct surrounding neighbourhood, bound by the green lines directly surrounding the school.

*Appendix D*

Scatter plot of the dependent variable, Ln(Sold Price) against the control variable 'adjusted IMD. The graph shows a positive linear relationship, as discussed in Section 3.4.



Appendix E

Correlation matrix between all variables.

	Sold Price	Ln(Sold Price)	No of Bedrooms	No of Bathrooms	Garage	Conservatory	Flat	Semi	Terraced	Detached	Good Condition	Bad Condition	Bespoke IMD	School Performance
Sold Price	1.00													
Ln(Sold Price)	0.93	1.00												
No of Bedrooms	0.57	0.58	1.00											
No of Bathrooms	0.51	0.48	0.57	1.00										
Garage	0.28	0.30	0.20	0.20	1.00									
Conservatory	0.21	0.24	0.22	0.23	0.10	1.00								
Flat	-0.17	-0.21	-0.36	-0.09	-0.08	-0.10	1.00							
Semi	0.03	0.09	0.13	-0.05	0.23	0.12	-0.21	1.00						
Terraced	-0.27	-0.31	-0.16	-0.16	-0.41	-0.20	-0.19	-0.68	1.00					
Detached	0.47	0.45	0.29	0.36	0.31	0.19	-0.09	-0.34	-0.31	1.00				
Good Condition	0.19	0.22	0.12	0.11	-0.03	0.00	-0.02	0.11	-0.10	0.01	1.00			
Bad Condition	-0.16	-0.20	0.00	-0.18	0.02	-0.04	-0.03	-0.05	-0.05	0.02	-0.22	1.00		
Adjusted IMD	0.51	0.63	0.22	0.17	0.22	0.12	-0.04	0.06	0.06	0.23	0.17	-0.14	1.00	
School Performance	0.42	0.44	0.19	0.13	0.13	-0.01	0.08	0.00	0.00	0.05	0.02	-0.07	0.43	1.00

Appendix F

Outlier Robust Regression results, to assess the impact of outliers on the model.

Independent Variable:	Coefficient
Terraced	0.0890
Semi	0.1723**
Detached	0.5146***
Num of Bedrooms	0.2048***
Num of Bathrooms	0.1213**
Garage	0.0178
Conservatory	0.0428
Good Condition	0.1209**
Bad Condition	-0.1260***
School Performance (*10 <sup>-3</sup> )	6.44
Adjusted IMD (*10 <sup>-3</sup> )	0.0253***
Constant/Intercept	10.865***

Note: The coefficient is significant at the \*10% level, \*\*5% level or \*\*\*1% using a two-sided test.

Source(s): Author's research, using STATA statistical software. See Section 3 for individual variable sources.