Econ H195B: Senior Honors Thesis

The Effect of School Quality on Prices versus Rents

Isaac Pariser

University of California, Berkeley

Advisor: Enrico Moretti

Abstract

It is a widely stylized fact that residential home values are higher in areas with good schools. Many researchers have studied the quantitative effects of school quality and sale prices, but very few have looked at the effect of school quality on rents. Specifically, there is little understanding on whether good schools are capitalized more in prices or rents. I build price-rent ratios for over 50,000 homes in Orange County, CA using a scrapped dataset from Zillow.com and compare the price-rent ratio of each home to the quality of its assigned schools. I find that prices capitalize school quality 2.8x more compared to rents. I find that this difference is largest in more affordable homes and in response to elementary schools. I include a number of additional data specifications to ensure the robustness of my results. I offer a possible explanation for the results and present possible extensions for future research.

1. Introduction

The price-rent ratio is a common metric used for comparing the value of a home on the purchase versus rental markets. Intuitively, the price-rent ratio measures how many years of rent need to be paid on a home before reaching its sale price It can be calculated by dividing the a home's purchase value by its annual rent. A high price-rent ratio indicates that prices are high relative to rents. This is a sign that demand for buying is high relative to the demand for renting. The opposite holds true for homes with low price-rent ratios.

There are many reasons that price-rent ratios can vary from city to city or even home to home. Increases in home values benefit homeowners and hurt renters, so expectations of future price increases in an area may drive up prices relative to rents. In addition to the purchase price of a home, homeowners bear the costs of property taxes and structural repairs, so homes with high added costs to ownership should have lower prices relative to rents. Lastly, homeowners and renters represent different subsets of the population, with different preferences, family profiles, and housing associated aspirations. These differences are reflected in the market prices of buying and renting. While these theories make reasonable sense, very few researchers have actually examined the scale to which home factors explain the differences in prices and rents in the data.

It is universally acknowledged that good schools are positively incorporated into the price of homes. Schools are clearly an important factor for homebuyers. Surveys have found that 91% of potential home buyers consider school zone boundaries in their search[1], 31% rank schools as their third most important consideration[2], and 25% report that schools were the deciding factor in their final decision[3]. Home prices reflect this

---

[1] Source: Realtor.com (2013)
[2] Source: Philadelphia City Planning Commission (2000)
[3] Source: National Association of Realtors (2017)

demand for school quality. A number of researchers have looked at the capitalization of school quality in home prices, and the consensus find that potential homebuyers are willing to pay 3-4% more for a one standard deviation increase in school quality.

However, the extent to which school quality is capitalized into residential rents has been overlooked. Specifically, there is very little research comparing the effects of school quality on prices versus rents. This is partially due to the difficulty of finding reliable rental data; unlike home sales, which get entered into county recorders' offices, there is no single, reliable repository for rental data. Still though, researchers should want to study the effects of school quality of price-rent ratios for the same reason they study the effects of school quality on prices. Understanding the downriver effects of increased school spending may help public planners and policy makers in state and local government make more informed decisions. Additionally, determining the capitalization of school quality on rents may help researchers more accurately gauge if allocations and investments in school quality are efficient. Lastly, studying the effects of school quality on prices versus rents may shine light on additional differences in preferences between homeowners and renters.

In this paper, I use a novel dataset scrapped from Zillow.com, an online real estate marketplace and database company to measure various effects of school quality on prices, rents, and price-rent ratios. My dataset lends itself naturally to creating high quality price-rent ratio estimates and features a number of qualities that set it apart from traditional rental data sources. I compare prices, rents, and price-rent ratios to a holistic measure of school quality from GreatSchools.com using a hedonic regression. I find that school quality has a much larger effect on prices than rents, with this differences existing across market segments and grade levels. I subject my results to a number of robustness checks to ensure validity, and offer potential explanations for this result. Overall, my findings shine light into a previously unexplored area of real-estate economics and lay a path towards future research in this topic.

2. Literature Review

According to the Tiebot Model (1956), consumers sort themselves into communities based on their preference patterns for local public goods (LPGs). Individuals with high willingness to pay for LPGs move into communities that provide their preferences for higher quality amenities. In turn, differences in desirability are incorporated into the costs of housing, either through local taxes or prices. Consumer willingness to pay for a public good such as school quality can therefore be inferred through housing markets.

There is a large body of economic literature which explores the effects of school quality on housing prices. The most frequent concern in these studies is the likelihood of omitted variable bias. Good schools tend to be located in areas with higher public amenities, lower crime, and higher quality homes. Since these variables are also positively incorporated into home prices, attempting to estimate the marginal willingness to pay for higher school quality without a strong control methodology will result in large upward biases. Black (1999) develops a model that is now employed by the majority of the literature. Black recognizes that there are a number of unobservable environmental and neighborhood factors which may influence the price of housing. To account for this, she selects homes that lie near the boundaries between school attendance zones. She then controls for neighborhood characteristics using fixed effects corresponding to each boundary. This design works on the principle that changes in school quality are expected to be discrete at boundaries, while changes in neighborhood characteristics are expected to be smooth. The difference in prices located on opposite sides of the school zone boundary line can therefore be explained by differences in school quality. Using this specification in a hedonic regression model, Black that a 5% increase in the average elementary school test scores for homes in the Boston is associated with a 2.1% increase in house prices. When comparing these results to a model without border fixed effects, she finds that failing to control for omitted-variable bias can result in estimates over 2x greater. Black's findings show that schools are significantly factored into housing prices, but failing to control for omitted variable bias may greatly overestimate the predicted effect.

A number of researchers (Bogart and Cromwell, 2000; Gibbons and Machin, 2003; Davidoff and Leigh, 2007; Bayer et al., 2007) have employed this model to replicate results in various other markets. Overall, the consensus of researchers estimate that a one standard deviation increase in average test scores results in a 3-4% increase in prices. Fack & Grenet (2008) hypothesize that in some areas, home prices may be less responsive to school quality if parents have the option to send their children to a nearby private school. They rank homes in Paris by their proximity to private schools and find that homes in the top quartile of private school proximity capitalize school quality into prices at roughly half the amount compared to homes in the bottom quartile of private school proximity. However, they do not examine the quality of private versus public schools in their dataset. The researchers recognize that comparisons between public and private schools cannot be made using their analysis and that results may differ in areas with similar quality of private versus public schools. Chiodo, Hernández-Murillo, & Owyang (2010) note that preferences for schools are heterogeneous and therefore prices may be non-linear with school quality. They include polynomial terms for test scores, finding that standard linear specifications may overestimate premiums at low levels of school quality and underestimate premiums at the high end.

Gibbons, Machin and Silva (2009) build on the econometric framework of the boundary discontinuity model by introducing a non-parametric discrete-cell matching design. Compared to the standard boundary discontinuity approach, this methodology allows for a fully non-parametric specification when controlling for housing characteristics. The researchers create matched pairs using property transactions from houses on opposite sides of school boundary lines. The homes in each matched pair have very similar housing attributes, but differ in school assignments. Since the homes in each pair are assumed to be identical in every regard but assigned school zones, the difference in price in each of the matched pairs can be attributed solely to the difference in school quality. The researchers demonstrate this model on an expansive set of property transactions in England and find effects of school quality on prices in line with the consensus. Furthermore, the researchers found prove results to be robust to a number of falsification tests in which they test if the model makes incorrect conclusions about matched pairs on imaginary school boundary lines.

These studies typically use "output based" measures of housing quality—test scores. Hanushek (1986, 1997) finds that previously used "input-based" measures of education quality, such as per-student spending and teacher-to-student ratio have little to no impact on student learning outcomes. Contemporary literature has recognized that test scores may also be a poor measure of school quality since students from privileged socio-economic backgrounds tend to have higher academic achievement. Some researchers have proposed using "value-added" metrics of school quality, measured by the improvement of one's standardized test performance over the course of a certain enrolment period. Brasington & Haurin (2006) question whether consumers have appropriate access to information about value-added metrics compared to output based metrics. They finds that there is a higher price premium for schools with good test scores compared to premiums for value added metrics, evidence that consumers form opinions on school quality using imperfect information.

While many researchers have studied school quality and house prices, very few researchers have examined whether school quality is capitalized more in housing rents or prices. In general, past studies comparing prices and rents have typically focused on the analysis of temporal variation of real-estate price and rent indexes and their macro-scale implications rather than the spatial distribution of prices versus rents. Case and Shiller (1990) are the first to suggest the study of price-rent ratios by claiming that high prices relative to rents may be a sign of market inefficiency. Himmelberg, Mayer, and Sinai (2005) observe that home prices are subject to speculative forces while rents tend to more accurately reflect the true value of housing. They hypothesize that departures from long-run price-rent ratios could be a sign that home prices are being supported by buyers speculating on future price appreciation rather than fundamental value. Gallin (2008) specifically examines the predictive power of the price-rent relation, building a time series model to examine cyclical patterns across prices and rents. He concludes that house prices tend to revert to historical price-rent ratios though it is difficult to predict when. After the run-up of price-rent ratios before the Subprime Mortgage Crisis, a number of researchers studied if high price-rent ratios are indicative of economic bubbles. Kishnor & Morley (2015) find that markets with high price-rent ratios are more sensitive to changes in mortgage rates and price expectations. Liu et al (2017) use price-rent ratios to determine correlations between current prices, expectations, and size of rational housing bubbles. Finally, Beracha and Johnson (2012) determine that profitable trading strategies based on price-rent ratios do exist, but only for those who are willing to change their tenure choice.

A number of researchers have sought to explain the variance in price-rent ratios. Cambell et al. (2009) perform a variance decomposition of the price-rent ratio using an autoregressive time-series model. They find that a majority of the change in price-rent ratios over time can be explained by changes in price, rent, and borrowing rate expectations. In recent years, a small number of researchers have begun to study cross-sectional differences in price-rent ratios. Pancak (2017) looks at aggregated price-rent ratio data across local municipalities in Connecticut and finds that local property taxes have an adverse effect on price-rent ratios while population age is positively correlated. Lee & Park (2017) use a Bayesian multi-level modeling approach to examine unit-level price-rent ratios in South Korea. They find evidence that housing characteristics such as size, floor number, and unit age are all significant in determining price-rent ratios.

Beracha & Hardin (2018) are evidently the first to look at the effects of school quality on prices versus rents. They use a standard hedonic regression model to compare the Florida Department of Education's A-F school ratings to premiums for buying versus renting. Like Black (1999), they use a border discontinuity approach to compare homes that lie on opposite sides of school assignment lines. Their results show a price premium of 4.3% and a rent premium of 2.6% generated by a one Florida letter rating increase. This put forward that schools are generally capitalized in home prices more than rents. They also find that schools have a larger effect on price-rent ratios in houses with family-oriented characteristics, suggesting that buyers with children place a higher premium for living near good schools compared to renters with children.

This paper makes at least three important contributions to the existing literature. First, this paper provides further evidence that school quality has a greater effect on home prices than rents. Second, this paper is the first to examine this result across grade levels and market segments. Third, this paper contributes to the literature by showcasing the use of new online data sources for housing prices and rents and encouraging their use for future economic analysis.

3. Data

Background

The price-rent ratio is commonly used when comparing prices and rents. Other metrics include rent-price ratios, CAP rates, and calculated buy versus rent premiums. Explicitly, I define the price-rent ratio is defined as the value of a home divided by its annual rent:

$$PR\ Ratio = \frac{Price}{12 * Monthly\ Rent}$$

By comparing a variable of interest to price-rent ratio, we can test whether it has a larger impact on prices or rents. Since prices are typically observed for owner-occupied homes and rents are only observed for rental properties, most researchers construct price-rent ratios using aggregated price and rent indexes. However, there are multiple issues with this approach. Size, quality of construction materials, and location greatly differ across homes that are renter versus owner-occupied (Hattapoglu & Hoxha, 2014). Thus, comparing aggregated prices and rents from distinct markets may not be a "like-to-like" comparison. Another potential problem is the modifiable areal unit problem (MAUP). MAUP occurs when continuous spatial data is aggregated by arbitrary or artificial boundaries. As a result, estimated relationships between variables at the aggregated level may differ and in extreme cases, even contradict the true relationship between variables at the granular level.

By constructing micro-level price-rent ratios, researchers can avoid the pitfalls of aggregation. Regrettably, property level price-rent ratios do not occur naturally; homes cannot be sold and rented simultaneously. As a result, researchers have designed imaginative ways to construct these ratios. Bracke (2015) uses properties in Central London that were involved in both rent and sale transactions within a short time frame. Chen & Ni (2010) develop a model of rent determination and impute unobserved rents for owner-occupied apartments across Shanghai and Shenzhen. Bram (2012) matches prices and rents of properties in luxury high rise Manhattan apartments in which renters and owners are likely to live next to one another in similar units. Yue Tan (2017) matches rental and sale listings from Craigslist by block, floor number, and unit area. In a much more rigorous use of matching, Lee & Park (2018) use price and rent contract data from the South Korean Ministry of Land Infrastructure and Transport to build price-rent ratios using sale and rental properties located on the same floor within the same apartment complex and contracted on the same month.

In this paper, I use a novel dataset of prices and rents from Zillow, an online real estate marketplace and database company. Zillow aggregates data from county, tax, building, and mortgage records to provide users access to information for homes of interest. In addition, homeowners are encouraged to claim their home on the site and upload additional information. Zillow uses this information to create estimates of potential sale and rental prices. While these estimates are not appraisals, they allow homeowners to approximate the potential value of their homes on the rental and sale markets. The details of the formula are proprietary, but Zillow states that estimates are first constructed using sales and rentals of comparable houses in a neighborhood then adjusted for property specific features using a rigorous machine learning algorithm.

Estimates are designed to be unbiased and are tested by Zillow to make sure they exhibit no serial patterns of over or underestimating. The accuracy of estimates often relies on the quantity and quality of the data available. Zillow provides the accuracy of their sale and rental estimates aggregated at the state, county, and major metropolitan level [Fig 1]. Accuracy can vary widely by state and county due to the availability of local records. For data not captured in local records, homeowners and landlords can upload information themselves. This is common when homeowners and landlords list their home on the Zillow marketplace, and areas with high volume housing markets tend to have higher accuracy scores.[4]

The Zillow price and rental estimates lend themselves naturally to building price-rent ratios. Matching prices and rents has already been handled by Zillow, and price-rent ratios can be calculated easily. Since price and rent estimates exist at the micro-scale,

[4] Unlike sale prices, there is no central repository for rent data. Thus, it is unclear what benchmark Zillow is comparing its rental estimates to. Zillow receives rental listing data from its online marketplace. Zillow also claims that it scrapes data from other rental listing sites, which could provide another measure to compare accuracy

we can be confident that we are not comparing two different stocks of housing of housing. Lastly, since all homes have estimates, there is less of concern of selection bias.

To measure the quality of assigned public schools, I use ratings from GreatSchools, an independent non-profit which aims to connect parents with information such as test scores, school demographics, teachers, and learning outcomes. Rather than rating schools solely on test scores, GreatSchools uses a holistic approach to rate schools on a 0 to 10 scale. GreatSchools builds its ratings using a weighted combination of the following pre-defined factors:

- *Test Scores*: a performance measure of school quality based on school's proficiency in state and local standardized testing. In California, this is largely through the California Assessment of Student Performance and Progress (CAASPP) System. Scores are compared to state averages across grades and subjects
- *Student Progress*: a value-added measure of school quality based on increases individual students have made on reading and math assessments during the past year or more. Student progress models are based on each state's own student growth model and compared to outcomes across the state
- *Equity*: a measure of a school's success in serving disadvantaged students and ensuring academic success for all students. A score is built by comparing a schools performance and value-added results across disadvantaged racial, ethnic, and socioeconomic groups to students throughout the state and within the same school
- *College Readiness*: a measure of school quality focused on assessing the degree to which a school prepares students for entrance into Post-Secondary education. This rating is based on High School graduation rates, college entrance exam scores, and Advanced Placement (AP) participation and

performance. In California, the percent of high school graduates who meet UC/CSU entrance requirements is also included.

- *Advanced Courses Ratings*: a measure of school quality focused on assessing the academic rigor of high schools based on student enrollment in advanced courses. This rating is determined by the number of advanced courses offered and enrolment in these courses relative to the state. A course is considered "Advanced" if it is part of the Advanced Placement (AP), International Baccalaureate (IB), or Middle Years Program (MYP) programs, or contains modifiers such as "Advanced", "Honors", or "Accelerated" in the course name.
- *Discipline and Attendance Issues*: used to identify schools with worrisome patterns of suspension and chronic absenteeism in their student body. Measured as out of out-of-school school suspension rates and proportion of students absent 15 days or more

Weights vary from school to school depending on the quality and availability of data. Circumstantially, elementary school ratings typically assign a .55 weight to test scores, a .3 weight to equity scores and the remaining .15 into value added metrics. For middle schools, ratings are split roughly .65 towards test scores and .35 towards equity scores. Highs schools incorporate additional measures, assigning college readiness a weight of .55 and advanced courses .1. Test scores and equity scores account for the remaining .2 and .1 weights. These ratings can be viewed as more accurate measure of school quality compared to test scores since they take into account many additional measures of educational success. In addition, these ratings are more likely to reflect consumer perceptions of school quality. Many home search websites such as Zillow provide information from GreatSchools. Hence, prices and rents may be more responsive to differences in GreatSchool ratings than compared to test scores alone.

Data Collection

The geographic area of focus for this study will be Orange County, California. Orange County ranks high in both price and rent estimate accuracy according to Zillow's metrics, with a median error of 3.2% for prices and 4.9% for rents. The makeup of real estate in Orange County is largely suburban and is dominated by Single Family Residences. This is important for multiple reasons. Firstly, Zillow does not estimate both prices and rents for condominiums, apartments, and multifamily homes, and these homes will not be included in this paper. By choosing a SFR dominant area, I decrease the possibility of selection bias and confounding market effects. Secondly, the Tiebot model is predicted to be more accurate in suburban markets[5]. These areas typically have a large number of individual communities, giving consumers a large set of choices for bundles of LPGs. Moving costs in suburban areas are also than urban and rural areas, removing frictions to efficient public good distribution. As a result of these factors, price and rent premiums in Orange County are more likely to reflect observed school quality.

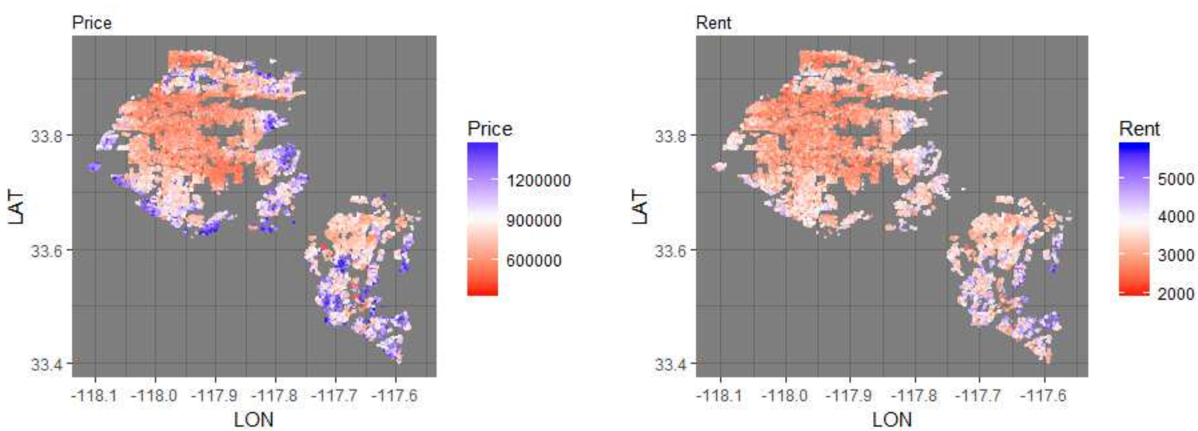
A cross section of price and rent estimates, property data, and school information were scrapped from the Zillow website using the rvest package in R. A simple random sample of 85,000 addresses were selected from openaddress.io, an open-source online address database. These addresses were used to create Zillow urls for each home. For each address in the list, a call was made to the Zillow website to extract the source html for the property in question. Websites often monitor web scrapping to block activity that may put excess strain on the website's servers, so a one-second time buffer was implemented between each call. Data was collected continuously during a 72 hour period in early April. A number of html and text parsing scripts were then used to extract home information and input into a dataframe for analysis. Data was cleaned to ensure accuracy [Table 2].

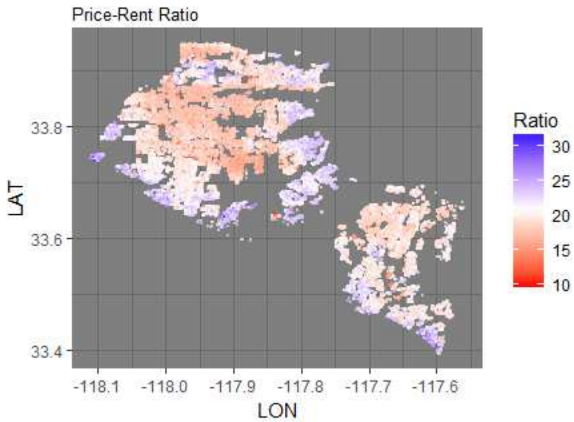
[5] Gruber (2016)

Since I am using price and rent estimates and not real transactional data, a number of alternate data specifications were created to test whether results are robust when excluding data predicted to be less accurate. One concern is that Zillow creates estimates for homes that no or very little transactional history. While price estimates can be inferred using home attributes, past sale prices represent the real market value of a home, including unobservable characteristics. It is fair to assume then that estimates for homes without recent transactional history are more likely to be inaccurate. To account for this, a subset of the data was created where only homes with a transaction within the last five years were included. A further concern is that rent estimates may be more inaccurate compared to prices. Unlike sale transactions, residential lease contracts are not publicly documented. Thus, rent estimates are especially subject to deviations from reality. I address this issue by creating a data subset of homes which were recently listed for rent on the Zillow marketplace. For this subsection, the Zillow rent estimates reflect actual transactional data. I identify these homes by checking if they include a description leftover from a recent rental listing and are currently renter occupied.

4. Summary Statistics

Fig. 1: Spatial Distribution of Data





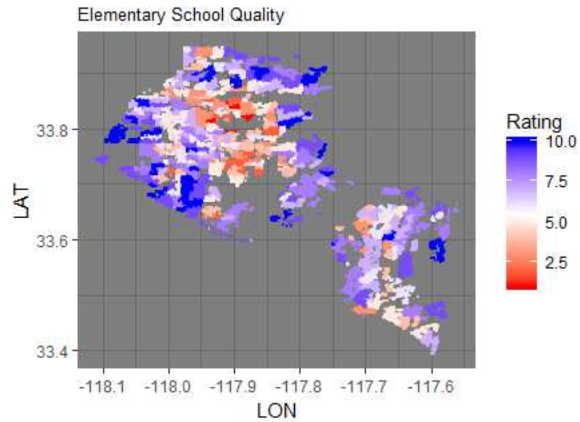


Table 3. Summary of Property Characteristics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
Price	448,500	639,400	748,500	803,200	898,300	3,110,000	232045
Rent	1,880	2,900	3,200	3,358	3,600	13,900	667.3
Price-Rent Ratio	12.14	18.08	19.32	19.69	20.89	44.77	2.44
Bedrooms	1.000	3.000	4.000	3.554	4.000	9.000	0.80
Bathrooms	0.500	2.000	2.000	2.364	3.000	6.500	0.69
Sqr. Ft.	823	1,410	1,781	1,915	2,291	5,075	662.7
Lot Size (ft)	1,386	5,600	6,534	6,966	7,700	26,570	2954.1
Age	1.00	41.00	51.00	49.14	60.00	118.00	16.4
Parking Spaces	1.00	2.000	2.000	2.046	2.000	10.000	0.78
Stories	1.00	1.000	1.000	1.432	2.000	6.000	0.54
Av. Distance to School (miles)	0.133	0.6667	0.9000	1.0450	1.3000	5.1670	0.577

Note:
Total Homes: 54,708
Owner Occupied: 45,920 (83.9%)
Renter Occupied: 8,788 (16.1%)
Claimed by Homeowner: 15,123 (27.6%)
Sold within Last 5 Years: 7,589 (13.9%)

Table 4. Summary of School Data

		School Rating		# Homes per School	
	# Schools	Mean	Std. Dev.	Mean.	Std. Dev.
Elementary	341	6.249	2.394	160.43	103.47
Middle	82	6.463	2.530	667.17	319.35
High	59	7.559	2.144	927.25	367.98

Fig 2. School Quality by Grade Level versus Price-Rent Ratio

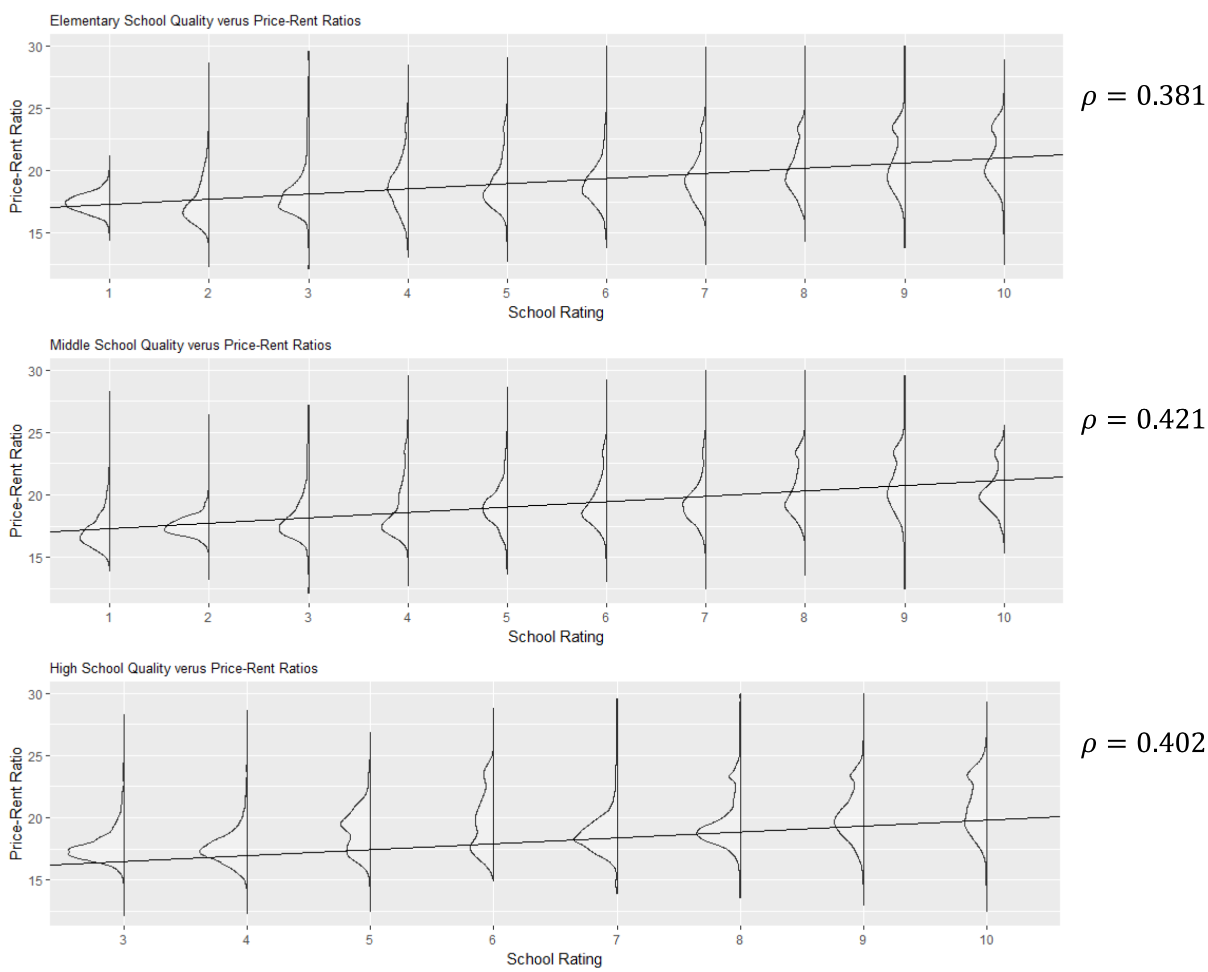


Table 5. Correlation of Variables

	Price	Rent	P-R Ratio	Owned	Bed	Bath	Sq Ft. (x100)	Lot Size (x100)	Age	Stories	Parking	Av. School Distance	E.S. Rating	M.S Rating	H.S Rating
Price	1	0.929	0.768	0.068	0.369	0.602	0.758	0.247	-0.450	0.396	0.215	0.226	0.378	0.424	0.428
Rent	0.929	1	0.489	0.098	0.453	0.686	0.826	0.223	-0.521	0.440	0.228	0.243	0.316	0.367	0.381
P-R Ratio	0.768	0.489	1	-0.007	0.123	0.274	0.406	0.205	-0.208	0.209	0.132	0.136	0.381	0.420	0.402
Owned	0.068	0.098	-0.007	1	0.100	0.106	0.129	0.049	-0.066	0.056	0.028	0.020	0.036	0.018	0.020
Bed	0.369	0.453	0.123	0.100	1	0.539	0.571	0.162	-0.246	0.314	0.128	-0.011	0.161	0.145	0.118
Bath	0.602	0.686	0.274	0.106	0.539	1	0.777	0.099	-0.561	0.501	0.230	0.179	0.245	0.269	0.277
Sq Ft. (x100)	0.758	0.826	0.406	0.129	0.571	0.777	1	0.268	-0.529	0.552	0.245	0.190	0.262	0.267	0.293
Lot Size (x100)	0.247	0.223	0.205	0.049	0.162	0.099	0.268	1	0.194	-0.116	0.127	-0.065	-0.020	-0.073	-0.059
Age	-0.450	-0.521	-0.208	-0.066	-0.246	-0.561	-0.529	0.194	1	-0.427	-0.184	-0.330	-0.320	-0.416	-0.436
Stories	0.396	0.440	0.209	0.056	0.314	0.501	0.552	-0.116	-0.427	1	0.118	0.141	0.195	0.226	0.245
Parking	0.215	0.228	0.132	0.028	0.128	0.230	0.245	0.127	-0.184	0.118	1	0.087	0.105	0.126	0.137
Av. School Distance	0.226	0.243	0.136	0.020	-0.011	0.179	0.190	-0.065	-0.330	0.141	0.087	1	0.105	0.159	0.242
E.S. Rating	0.378	0.316	0.381	0.036	0.161	0.245	0.262	-0.020	-0.320	0.195	0.105	0.105	1	0.692	0.580
M.S Rating	0.424	0.367	0.420	0.018	0.145	0.269	0.267	-0.073	-0.416	0.226	0.126	0.159	0.692	1	0.787
H.S Rating	0.428	0.381	0.402	0.020	0.118	0.277	0.293	-0.059	-0.436	0.245	0.137	0.242	0.580	0.787	1

Exploratory Analysis

We observe that large differences exist in prices, rent, and price-rent ratios across Orange County and both the County and localized level. Differences at the city level may be due to differences in property tax rates and local laws, while differences at the local level must be due to more home specific factors. Orange County can be split into North and South with both regions displaying variation in home values and school quality.

At first look from the maps in Fig. 1, areas with good schools seem to overlay with areas with high prices and high price-rent ratios. Fig 2 shows that there is a general correlation between school ratings and price-rent ratios. Prices seem to incorporate school quality more than rents. Significant multi-collinear effects exist though across home attributes and school quality though. School ratings and price-rent ratios both increase with home features. As a result, we are not yet certain if the relationship between school quality and prices versus rents is causal, or due to other factors. In general, the high multi-collinearity between variables will make causal arguments less interpretable. For example, predicting the effect of adding an additional bedroom while keeping all other variables constant is not intuitive since home size is also expected to increase.

For both prices and rents, the mean is greater than the median. This means that prices and rents are right skewed. Home values have a hard lower bound but no upper bound, so extreme values are more likely at the high end of the market. Price-rent ratios are also skewed right but by a smaller margin, indicating that they more closely follow a normal distribution.

5. Model

Housing Characteristics and Average School Quality

Throughout this paper, I use a standard hedonic regression model[6]. The hedonic regression is a revealed preference model for decomposing the price of housing into its constituent characteristics. Each coefficient in the regression can be viewed as consumer's willingness to pay for a certain feature. Prices and rents are modeled as a linear combination of coefficients and attributes:

$$\log(Price_{ian}) = \alpha + \theta X_{ian} + \beta av.school_a + \phi FE_n + \epsilon_{ian}$$

$$\log(Rent_{ian}) = \alpha + \theta X_{ian} + \beta av.school_a + \phi FE_n + \epsilon_{ian}$$

$$PR_{ian} = \alpha + \theta X_{ian} + \beta av.school_a + \phi FE_n + \epsilon_{ian}$$

[6] For a complete overview of the theory, design, and application of linear modelling, see Fox (1997)

$Price_{ian}$, $Rent_{ian}$, and PR_{ian} are the price, rent, and price-rent ratio estimates for house i in school assignment a in neighborhood n. X_{ian} is a vector of housing characteristics that includes variables for number of bedrooms, number of bathrooms, square footage, lot size, age, and number of floors, number of parking spaces, and average school distance. Polynomial terms are included for square footage, lot size, and age. $av.school_a$ is the average school rating across assigned elementary, middle, and high schools. FE_n is a vector of neighborhood fixed effect dummy variables.

Price and rents follow a log normal distribution. As standard in hedonic regression models, prices and rents are log transformed to impose linearity between the dependent variable and the regressors. The coefficients can be interpreted as measuring the percent change in prices and rents caused by unit changes housing characteristics. Specifically, for coefficient β, a unit change in an explanatory variable corresponds to a 100β% change in log prices or log rents. Price-rent ratios are not transformed, and a 1 unit change in an explanatory variable correlates to a β change in price-rent ratios.

There are a number of notable assumptions used in the linear regression model. Explanatory variables are assumed to be uncorrelated (no multi-collinearity) and the variance of errors is assumed to be constant across observations (no heteroscedasticity). We already know that the multicollinearity assumption is violated as housing attributes such as lot size, number of rooms, and square footage are highly correlated [Table 5]. The homoscedasticity assumption is typically broken if the number of observations falls at tails of the data. The violation of either assumption results in higher standard errors, but no bias.

An assumption of more critical importance to the model is that the errors terms are uncorrelated with the explanatory variables (exogeneity). Violations of this assumption will result in biases in the coefficients. Endogeneity typically results through omitted variable bias. Areas with good schools also tend to have other high quality public amenities, may be situated in more attractive geographic settings, and could higher quality houses with luxury features. If these factors are not accounted for in the model, the model will overstate the effect of school quality and the school rating coefficient will be skewed upward. I control for omitted variables by only comparing homes assigned to different schools in the same neighborhood. I do so through neighborhood fixed effects. Neighborhood boundaries are drawn using local and natural features, so it would be fair to assume that the distribution of amenities follows neighborhood lines. Still though, there may be sources of omitted variable bias. Firstly, it is possible that home quality or the distribution of amenities is not continuous at school zone boundaries. Consumers who prefer good schools could also prefer luxury home features, so non-observed factors may change within a neighborhood as soon as school boundary lines are crossed. To test for endogeneity, I use a Pearson correlation test. The test rejects the null if the correlation between the error terms and school quality is significantly different than zero. I also compare my estimated coefficients to that of the literature to determine if my results deviate from expected ranges.

One potential issue with this design is overlap between neighborhood fixed effects and school zone boundaries. Using neighborhood fixed effects, the model identifies the effect of changes in school ratings *within* neighborhoods. However, it is possible that all the homes within a neighborhood are assigned to the same schools. Zero variation within fixed effects may have two effects on the regression output. Most likely, standard errors will be higher. Neighborhoods in which all homes are assigned to the same school are not included in the calculation of the coefficient for school quality. This essentially shrinks the sample size, resulting in higher standard errors. Similarly, if neighborhoods are removed from the calculation, there may selection bias. This would occur if the neighborhoods with no school variability respond differently to changes in school quality compared to neighborhoods with school variability. There is no intuitive reason to believe this is the case, so coefficients are expected to be unbiased in this regard. To examine the effects of overlap between school assignments and neighborhood boundaries, I calculate alternate specifications in which I swap neighborhood fixed effects for city fixed effects and remove geographic fixed effects altogether. These results are presented in Table 6 columns 1-2, 4-5, and 7-8. If standard errors spike as neighborhood fixed effects are included, this is a sign of a lack of school variation within neighborhoods.

Non-linear Effects of School Quality

Researchers have observed that prices may react non-linearly with school quality. Therefore, it would be worthwhile to test whether this also holds true for rents, and whether there is a significant difference in the convexity of prices versus rents. If price-rent ratios respond non-linearly with changes in school quality, there is evidence that prices and rents have different convexities with school quality. I run the same regressions above, but now include quadratic terms for school quality:

$$\log(Price_{ian}) = \alpha + \theta X_{ian} + \beta_1 av.school_a + \beta_2 av.school_a^2 + \phi FE_n + \epsilon_{ian}$$

$$\log(Rent_{ian}) = \alpha + \theta X_{ian} + \beta_1 av.school_a + \beta_2 av.school_a^2 + \phi FE_n + \epsilon_{ian}$$

$$PR_{ian} = \alpha + \theta X_{ian} + \beta_1 av.school_a + \beta_2 av.school_a^2 + \phi FE_n + \epsilon_{ian}$$

The effects of average school quality on price-rent ratios are determined to be nonlinear if the coefficient for the quadratic term is found to be statistically significant. If price-rent ratios are increasing and convex with school quality (quadratic term is positive), prices are especially responsive to increases in school quality at the upper end compared to rents. If price-rent ratios are increasing and concave with school quality (quadratic term is negative), prices are especially responsive to changes in bad schools compared to rents.

Effects across Market Segments

We may also be interested if price-rent ratios respond differently to school quality at various levels of price. The demographics of buyers and renters are different across low versus high value homes, so it is likely that the capitalization of school quality on prices and rents changes across market segments. I segment homes based on their price quartile within each neighborhood and regress each separately:

$$PR_{ian}^{q} = \alpha + \theta X_{ian}^{q} + \beta av.school_{a}^{q} + \phi FE_{n}^{q} + \epsilon_{ian}$$

Where q indicates price quartile. For segments where the coefficient for school rating is high, the difference in school quality capitalization between prices and rents is greater.

Effects of Individual Grade Levels

Up to this point, we have been looking primarily at the effects of *average* school quality. Consumers may rank elementary, middle, and high school quality at different levels of importance though. I now look at individual school ratings to compare differences in the capitalization elementary, middle, and high school quality on prices and rents. Due to correlations among elementary, middle, and high school ratings, school ratings for a certain grade level cannot be regressed alone without risk of capturing the effects of other schools. One way to control for this is to regress on elementary, middle, and high school ratings simultaneously:

$$\log(Price_{ian}) = \alpha + \theta X_{ian} + \beta_1 e.school_a + \beta_2 m.school_a + \beta_3 h.school_a + \phi FE_n + \epsilon_{ian}$$

$$\log(Rent_{ian}) = \alpha + \theta X_{ian} + \beta_1 e.school_a + \beta_2 m.school_a + \beta_3 h.school_a + \phi FE_n + \epsilon_{ian}$$

$$PR_{ian} = \alpha + \theta X_{ian} + \beta_1 e.school_a + \beta_2 m.school_a + \beta_3 h.school_a + \phi FE_n + \epsilon_{ian}$$

A large issue with this model is that it does not account for differences in school assignments not included in school ratings or preferences that do not lie in line with school ratings. For example, take two identical neighboring homes in school assignment zone A versus B. Both assignment zones have identically rated elementary and middle schools, but zone A's high school is rated 6 and zone B's high school is rated 7. Since all other schools are equally ranked, the model predicts that the entirety of the difference in prices is due solely to the difference in high schools. It is possible though that consumers have a slight preference for the middle school in assignment zone B. In this case, the effect of high school quality is overestimated as part of the price difference should be attributed to the effects of middle schools. To avoid capturing the effects of other grade levels, a better measure would be to compare homes in the same neighborhood that are assigned to the same elementary and middle schools, but different high schools:

$$\log(Price_{ian}) = \alpha + \theta X_{ian} + \beta e.school_a + \phi FE_n + \delta_m FE_{m.school} + \delta_h FE_{h.school} + \epsilon_{ian}$$

$$\log(Rent_{ian}) = \alpha + \theta X_{ian} + \beta e.school_a + \phi FE_n + \delta_m FE_{m.school} + \delta_h FE_{h.school} + \epsilon_{ian}$$

$$PR_{ian} = \alpha + \theta X_{ian} + \beta e.school_a + \phi FE_n + \delta_m FE_{m.school} + \delta_h FE_{h.school} + \epsilon_{ian}$$

$FE_{m.school}$ and $FE_{h.school}$ are vectors of fixed effects corresponding to middle school and high school assignments. The differences in the dependent variables due to high school ratings can be attributed solely to the difference in high school assignments. This regression design is replicated to estimate the effects of elementary and middle schools, using fixed effects of school assignments for corresponding grade levels. Coefficients for individual schools are expected to be smaller compared to changes in the average across all three grade level.

School Quality and Tenure Choice

Lastly, I examine whether school quality effects tenure choice. Typically, a generalized linear model would be best for predicting the probably of a binary variable as it ensures that the predicted probabilities are bounded between 0 and 1:

$$Owned_{ian} = f(\alpha + PR_{ian} + \theta X_{ian} + \beta av.school_a + \phi FE_n + \epsilon_{ian})$$

$Owned_{ian}$ is related to the linear combination of the independent variables via a link function, $f()$. Logit models use the CDF of the log distribution and map the linear combination of explanatory variables to the log odds of the dependent variable.

Unfortunately, generalized linear models do not lend themselves well to interpreting the effects of individual coefficients since the marginal effect of one explanatory variable is dependent on the values of the others. As a result, I opt to use a simpler linear probability model:

$$Owned_{ian} = \alpha + PR_{ian} + \theta X_{ian} + \beta av.school_{a} + \phi FE_{n} + \epsilon_{ian}$$

A one unit increase in average school quality corresponds to a β unit increase in the probability of ownership. School quality is determined to have no effect on tenure choice if the coefficient for school quality is not significant. The results of the logit model are included for comparison.

6. **Robustness Checks**

Biased Measurement Error

An important consideration in my analysis is that I am using estimates of prices and rents in place of actual home transactions. In reality, there is likely some unknown measurement error between the price and rent estimates and actual market values. A high degree of measurement error will cause the models to significantly underreport standard errors. Zillow claims that estimates are unbiased, but if measurement error is correlated with the independent variables in the model, coefficients will be biased. I test if my results are robust by comparing them to alternative data specifications where estimates are less likely to deviate from true values. I replicate results from regressions 1-3 using the two alternate data specifications outlined earlier: one limited to homes sold in the within the last five years (spec 1) and one limited to homes previously listed for rent on Zillow (spec 2). I create two tests—one where I test for statistically significant differences between two coefficients:

$H_0 \colon \beta_1 = \beta_2$

H_A: $\beta_1 \neq \beta_2$

And one in which I test for statistically significant equivalence between two coefficients:

H_0: $\beta_1 \neq \beta_2$

H_A: $\beta_1 = \beta_2$

I calculate p-values using the methodology specified by Clogg, Petkova, & Adamantios (1995) to examine if the coefficients calculated using the full dataset are significantly different than those calculated using the alternate specifications. I calculate the following z-score corresponding to the number of standard deviations the difference between coefficients is from zero:

$$Z = \frac{\beta_{Full\ Data} - \beta_{Alt.Spec}}{\sqrt{(SE\beta_{Full\ Data})^2 + \left(SE\beta_{Alt.Spec}\right)^2}}$$

Where $SE\beta$ is the standard error of β. At a statistical significance of 10%, the test for difference rejects the null for values of $|Z| > 1.28$ while the test for equivalence rejects the null for values of $|Z| < .12$.

Goodness of Fit and Model Diagnosis

The R^2 value corresponds to the proportion of the variance in the dependent variable predicted by the independent variables in the model. For example, an R^2 of .7 for regression 1 indicates that 70% of the variance in log prices is explained by the independent variables in the model. The remaining 30% can be attributed to omitted variables or inherent noise in prices. While this is a simple way to test for goodness of fit, a low R^2 does not necessarily indicate that the model is misspecified. To some extent, prices and rents are set by random factors which impose some noise on the data. Thus, there is some variance in home values that is expected to be unexplained by the model. This holds especially true for price-rent ratios, which are subject to noise from both prices and rents.

Residual plots are included in the appendix for a graphical analysis of additional model diagnostics. Residuals should be randomly dispersed across predicted values. If the residuals do not appear normally distributed around zero for a given predicted value, the model may be misspecified. If the variance of residuals is not constant across predicted values, there is likely heteroskedasticity. As specified earlier, bias can arise in the coefficients when error terms are correlated with independent variables. I test for endogeneity using the Pearson's correlation coefficient, which follows a t-distribution with n-2 degrees of freedoms. Estimates are concluded to be exogenous if the correlation between independent variables and residuals are not significantly different than zero.

7. Results

Physical Housing Characteristics

As expected, both prices and rents increase with building area and are slightly concave, indicating that the marginal utility of space is decreasing. While prices capitalize changes in square footage 25-32% more than rents depending on size, rents are much more responsive to additions to the number of bedrooms and bathrooms. Rents may be particularly sensitive to changes in bedroom and bathroom counts because renters could be looking for homes that fit a specific dwelling arrangement or required room threshold. Potential homeowners generally value overall size over room count because they have the option to remodel to fit their preferences. The result is that prices are more responsive than rents to the overall square footage and lot size. In general, the model predicts that for the average 1,915 sq. ft. Orange County home, adding an additional 12 x 12 bedroom while keeping lot size fixed results in an average 3.13% increase in prices, a 3.60% increase in rents, and a -.093 decrease in price-rent ratios. Price-rent ratios decrease by -.232 and -.135 for additions to bedrooms and bathrooms respectively holding all other variables constant, and are increasing and convex with square footage.

Price-rent ratios are convex with age and either increase or decrease depending on the year built. For new homes, prices drop very quickly compared to rents. The model predicts that prices decrease by -4.01% while rents decrease by -1.08% over the first ten years of a home's lifespan. Price-rent ratios for homes built before 2000 are nearly a full unit lower than for those built in 2019. Buyers pay a premium compared to renters for new homes because they bear the cost of home upkeep and repair while renters typically levy these costs onto the landlord. As homes gain more wear-and-tear, the risk of structural problems increases, and potential buyers may expect more out of pocket repair costs. Interestingly, the model predicts that price-rent ratios begin to increase with age for homes built before 1972. One reason for this result may be that as homes age past this point, they are more likely to have undergone massive renovations or repairs that put their structural risk on level with that of newer homes. These older homes thus pose the same out of pocket repair risks for potential buyers and the advantage to renting over buying lessens.

There is no significant effect of building stories and number of parking spaces on price-rent ratios. The addition of a second story has an equal -.9% effect on both prices and rents which cancel each other out in price-rent ratios. This indicates that both renters and buyers have similar disutility for this feature. Parking spaces, on the other hand, do not seem to have much of an effect on either prices or rents. While this result contradicts expectations, it may reflect the suburban nature of Orange County. Most homes in the sample have access to free public parking on their streets, so additional on property parking spaces may be inconsequential. While there could be more incite through distinguishing garage spaces from driveway spaces, there was no data.

Average School Quality

Both prices and rents increase with school quality. A one unit increase in average school ratings is associated with a 1.7% increase in prices and .6% increase in rents compared to homes in the same neighborhood. A one standard deviation increase in average school ratings corresponds to a 3.34% increase in prices and a 1.18% increase in rents. These estimates are consistent with expectations from the literature. Applying these estimates to the average price and rent estimates from Table _ implies a price premium of $26,844 and an annual rent premium of $475.30 for a one standard deviation increase.

There is a 1.1% difference in price versus rent capitalization, with prices 2.8x more sensitive to changes in average school quality. A one rating increase corresponds to a .231 unit increase in price-rent ratios, and a one standard deviation increase in average school ratings corresponds to a .454 price-rent ratio increase. In other words, homeowners place a value of roughly 5.5 months of rent on their homes for a one standard deviation increase in average schools.

Comparing the coefficients obtained using neighborhood fixed effects versus a city fixed effects model, we see that the coefficient for average school rating decreases as more localized fixed effects are added. This is a sign that the model is controlling for additional omitted variables positively correlated with both school rating and home value. Since the average school rating coefficient also decreases for price-rent ratios, we can infer that the omitted variables at the city and county level are more highly correlated with prices than rents. Examining the standard errors, there do not seem be large spikes as city or neighborhood fixed effects are added. This indicates that there is sufficient variation in average school ratings within neighborhoods.

Non-linearity in Price-Rent Ratios

Both prices and rents are convex with school quality. A graphical representation is presented in Fig _. A linear model would over predict the effects of school quality in areas with low-rated schools and under predict the effects of school quality in areas with highly-rated schools. There is a 1.35% price increase between home rated 4 versus 5 while prices increase by 2.08% across homes with ratings of 8 versus 9. For rents, these differences in ratings equate to .22% and .97%. Overall, the non-linear effects in prices and rents nearly cancel out. The model finds no significant non-linear effects between average school ratings and price-rent ratios.

Differences among High versus Low Value Homes

Prices show the largest sensitivity to school changes for low value homes, with a one rating increase corresponding to a 1.1%, .50%, .67%, and .31% increase in prices across quartiles. These percentages equate to $7241, $3693, $5498, and $3062 in nominal terms for mean prices in in quartile. Notably, prices have the largest percentage and nominal response to school quality for the lowest quartile, and a very small percentage response in the highest quartile. Rents change by .40%, .27%, .46%, and .32% across quartiles. In nominal terms for mean rents in each quartile, these percentages equate to $144, $96, $192, and $144 in annual rent. There does not seem to be a pattern in school capitalization across quartiles for rents. As home values increase, prices respond increasingly less to school quality compared to rents. A one standard deviation change in school quality corresponds to a .261, .081, and .061 increase in price-rent ratios across quartiles, with an insignificant change in the fourth quartile.

Price Rent Effects of Different Grade Levels

For homes in the same neighborhood and differing in only one school assignment, prices are predicted to increase by .63%, .36%, and .21% in homes assigned to 1 rating higher elementary, middle, and high school ratings, respectively. Rents are only significantly predicted to increase for elementary schools by .25%. Price-rent ratios increase .080, .047, and .033, significant at a p-value of .1. A one standard deviation increase in elementary, middle, and high school ratings, respectively, is predicted to cause a 1.42%, .85%, and .42% increase in prices, a .57%, .41%, and .10% increase in rents, and 0.18, 0.11, and 0.07 increase in price-rent ratios. Elementary schools have a 3x greater price premium compared to high schools while middle schools have a 1.7x greater capitalization rate. Between prices and rents, elementary schools are capitalized 2.5x more in prices, and though the rent estimates where not statistically significant, middle schools and high schools are capitalized 2.1x and 4.3x more in prices.

From earlier results, there is evidence that rents are generally less responsive to increases in school quality compared to prices. This means that there already less of a measurable difference from zero for school estimates for rents. This is compounded by the restrictive fixed effects model, which imposes higher standard errors on the coefficients. As a result, the coefficients for rents were not found to be significant. When compared to the model in which elementary, middle, and high school rating where regressed simultaneously, it is clear why this methodology is needed though. The simultaneous regression model overstates the effects of middle school quality and understates the effects of high school quality. This is evident in the estimated effect of high school rating on rents, where rents are expected to decrease -.1% for a one rating increase in high school ratings. Intuitively, there is no logical reasoning for why renters should prefer worse high schools, so this is clear evidence that simultaneous regression model is either misspecified or has omitted variable bias. While the school fixed effects model has lower significance, we can be more confident in its coefficients.

School Quality and Homeownership Rates

Despite the links between prices, rents, price-rent ratios, and average school quality detailed above, increases in school quality have no significant effects on homeownership rates. Holding all variables constant, homeownership decreases by 10.0% for every unit increase in price-rent ratios. We can infer that while schools have a positive effect on price-rent ratios, changes in homeownership rates due to price-rent ratios are not due to the differences in school quality.

Robustness

Examining the results of the alternative specifications, we see no significant difference between the robustness checks and the full data specification. For specifications 1 and 2 respectively, we achieve p-values of .293 and .329 for prices, .291 and .333 for rents, and .292 and .337 for price-rent ratios. While these values indicate that the estimates from the two specifications do not significantly differ from the full dataset, the results are not close enough for statistical equivalence. The alternate specifications calculated slightly larger coefficients, but this may equally be due to selection bias rather than serial measurement errors in the Zillow estimates.

A visual analysis of the residuals plots for the regressions on prices, rents, and price-rent ratios reveals that variances are roughly constant, indicating that the assumption of homoskedacity is roughly achieved. The model does not over or under predict at different levels of fitted values, indicating that there is no gross misspecification, and likely no endogeneity. Testing for exogeneity with school ratings, there is a less than .001% chance that the residuals and average school quality are correlated. Strangely though, there are noticeable diagonal patterns in the residual plots, especially for log rents. This means that at cross sections of fitted values, error terms follow discrete patterns. While this doesn't mean anything necessarily wrong with the results, it may suggest that there are undiscovered artifacts in the Zillow estimates. This result could occur because a number of the variables in the model are discretely distributed.

8. Discussion

In summary, the results indicate that prices are universally more responsive to changes in school quality compared to rents. This difference is largest when looking at the effects of elementary schools and in the affordable market segment. I present two possible explanations for these result. One, buyers may place a larger premium on school quality compared to renters because they have a higher marginal utility for education. Two, consumers are not responding to school quality, but to the expectation that future price increases may be higher for homes in good school assignment zones.

It is difficult to gauge the validity of the latter explanation. Home values are expected to respond positively if the quality of their assigned schools increase. If prices are in equilibrium with school quality, school quality should not be an accurate predictor of future price changes. Areas with good schools should already have high prices, so prices should not be expected to increase more. Consumers may base their decisions on irrational expectations or incomplete information, on a shared belief that homes with good schools will appreciate more. There are likely many dynamics between home prices, school quality, and expectations that unfortunately cannot be measured in a cross sectional dataset.

There is more substantial evidence supporting the theory that consumers are indeed responding to school quality based on their preferences for education. Demand for school quality is correlated with demographic and family factors which vary across homeowners and renters. To begin with, in Orange County, the 2017 American Community Survey finds that 1.3% more homeowners have children compared to renters. Prices therefore may incorporate school quality more than rents simply because school quality is a non-issue for more renters than homeowners. There are multiple factors pushing families with children towards homeownership. Buying a home requires a level of long-term commitment, a consideration that is less likely to discourage families with children. Homeownership is also viewed as part of the “American Dream” and there are strong cultural beliefs that homeownership is more beneficial for raising a family. We see evidence of this in the large effect of school quality on prices within the lower quartile, possibly driven by an influx of first-time buyers expecting to start a family. For these consumers, school quality would be a critical factor. Renters in the bottom quartile show no notable affinity towards school quality. Buying a home, even an affordable one, requires some level of income and financial stability. Younger consumers, less likely to have children, are pushed towards rental markets. As we look towards more expensive market segments, it is likely that the demographics of homeowners and renters are more similar. As a result, prices and rents have rates of school capitalization more in line with one another.

Using the Thibaut model, this framework does well for explaining why both prices and rents could be convex with school quality. Consumers without children are more likely to sort themselves into communities with poor school quality and avoid paying a school quality premium. Consumers with children are more likely to sort into communities with good schools, substituting affordability with school quality. Thus, price and rent responses are much larger within communities with good schools since more consumers have a high willingness to pay for an additional unit of school quality. The opposite may hold true in communities with bad schools, where more consumers are indifferent to differences in school quality. The finding that price-rent ratios are linear with school quality indicates that while child rates may still be greater in homeowners versus renters, this sorting behavior is similar across renters and buyers.

Elementary school are generally acknowledged to be the most critical point in a child's educational development. Differences between buyers and renters are evident here, where buyers pay much higher premiums for good elementary schools. Both buyers and renters have muted responses to increases in high school quality. Consumers instead put greater weight in factors such as athletics or arts programs which are not captured in school academic ratings. Though estimates for rents are not statistically significant, it is worth noting that the effect of school quality on price-rent ratios is lower for older grade levels. If we expect renters to be younger on average compared to buyers, they should be increasingly less responsive to older grade levels compared to buyers. The results suggest that price-rent ratio capitalization is determined primarily by the price component.

Even among homeowners and renters with children, there are additional reasons to believe that homeowners have higher preferences for school quality. According to the 2017 ACS, in Orange County, high school graduation rates are 3.0% higher for homeowners, college attendance is 10.0% greater for homeowners, and homeowners are over twice as likely to have attended graduate school. Demand for school quality is positively associated with income and educational attainment, both which are factors correlated with homeownership. Parents with a high level of educational attainment likely have high preferences for the school quality of their children. Premiums for school quality may be reflected more in prices than rents for this additional reason.

9. Conclusion

In this paper, I constructed a novel dataset of home-level using data scrapped from Zillow.com to analyze the relationship between assigned school quality and prices versus rents. I find that school quality has a significantly positive effect on both prices, rents, and price-rent ratios, indicating that school capitalization is greater in prices than rents. I test these results for non-linear effects and find that while prices and rents are convex with school quality, price-rent ratios are not. Prices have much larger capitalization rents in elementary schools and affordable homes, both segments with large price premiums. This suggests that the price-rent ratio responses to school quality are dependent on the price component. Tenure choice does not change with school quality though. I subject my results to a number of robustness checks to test their validity.

There are a number of extensions of this research. First, my results should be replicated in different markets to examine external validity. Very few researchers have looked at the effects of school quality or prices versus rents, and establishing best practices ad consensus estimates will help legitimize these studies for use in public policy and decision making. Different methodologies, assumptions, and econometric designs should be used on the data to ensure reproducibility and robustness of the results. Second, the analysis would greatly benefit greatly from the introduction of a time component. Using a cross-sectional dataset for studying home values is an inherently problematic approach. Prices and rents are set through dynamic processes, and a cross sectional dataset fails to capture intricate relationships between variables in the model. Home prices are sticky, meaning that there is a lag between changes in school quality and changes in prices. As a result, current home values may not have fully incorporated current school ratings. These problems could be addressed using a panel dataset. Future researcher looking to use Zillow estimates should scrape for a continuous period to build history in the data. Lastly, future research should also test the theory that differences in family composition are the root of differences in price and rent capitalization. While household level family data would be difficult to match with specific home level price and rent estimates, aggregated data offers a good starting point.

10. Appendix

Table 1: Zillow Estimate Accuracy									
		Price				Rent			
	N	Within 5% of Actual	10%	20%	Median Error	Within 5% of Actual	10%	20%	Median Error
National	135.3M	50.2%	71.5%	85.1%	5.0%	39.0%	61.0%	82.9%	7.1%
California	10.0M	55.3%	76.8%	89.5%	4.3%	38.6%	61.0%	82.7%	7.1%
Orange County	834.9K	64.6%	83.8%	93.6%	3.2%	50.7%	73.7%	89.7%	4.9%

Table 2: Data Cleaning

Filtering Condition	**Rows Removed**	**N**	**Mean Price-Rent Ratio**
No filter	0	85,000	20.47
Address is non-residential	13,199	71,801	20.47
Address is non-SFR	3,662	68,139	20.47
Missing housing attributes	8,798	59,341	19.86
Remove outliers; trim top and bottom 1% of prices and rents	3,418	559,23	19.91
Remove sparsely populated neighborhoods (less than 10 homes)	1,106	54,817	19.69
Final Dataset	**30,346**	**54,654**	**19.69**

Table 6. Housing Attributes on Prices, Rents, and Price-Rent Ratios Across Fixed Effects

	Dependent variable:								
	log(Price)			log(Rent)			Price-Rent Ratio		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
# Bedrooms	-0.019***	-0.009***	**0.0002**	0.001*	0.009***	**0.012*****	-0.412***	-0.347***	**-0.232*****
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0004)	(0.014)	(0.011)	(0.010)
# Bathrooms	0.014***	0.019***	**0.018*****	0.023***	0.024***	**0.024*****	-0.194***	-0.126***	**-0.135*****
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.021)	(0.017)	(0.015)
Sqr. Ft. (100 ft.)	0.034***	0.026***	**0.022*****	0.024***	0.021***	**0.019*****	0.175***	0.084***	**0.031*****
	(0.001)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.007)	(0.006)	(0.005)
Sqr. Ft.^{2} (100 ft.)	-0.0002***	-0.0001***	**-0.00002*****	-0.0001***	-0.0001***	**-0.0001*****	-0.0005***	0.001***	**0.002*****
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00000)	(0.00000)	(0.0001)	(0.0001)	(0.0001)
Lot Size (100 ft.)	0.001***	0.003***	**0.003*****	0.001***	0.001***	**0.002*****	0.003***	0.024***	**0.035*****
	(0.0001)	(0.00005)	(0.00004)	(0.00004)	(0.00003)	(0.00003)	(0.001)	(0.001)	(0.001)
Lot Size^2 (100 ft.)	-0.00000***	-0.00001***	**-0.00001*****	-0.00000***	-0.00000***	**-0.00000*****	0.00003***	-0.00002***	**-0.0001*****
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Age	-0.002***	-0.003***	**-0.004*****	-0.001***	-0.001***	**-0.001*****	-0.022***	-0.045***	**-0.062*****
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.002)	(0.002)	(0.002)
Age^2	0.00002***	0.00003***	**0.00003*****	0.00000	0.00000***	**0.00000*****	0.0004***	0.001***	**0.001*****
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00002)	(0.00002)	(0.00002)
# Stories	-0.014***	-0.009***	**-0.009*****	-0.012***	-0.010***	**-0.009*****	-0.039*	0.023	**0.004**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.020)	(0.016)	(0.015)
#Parking Spaces	0.003***	-0.0005	**0.0003**	0.001**	-0.001	**0.0001**	0.020*	-0.003	**0.005**
	(0.001)	(0.001)	(0.0005)	(0.001)	(0.0004)	(0.0004)	(0.011)	(0.009)	(0.008)
Av. Distance to School	0.023***	0.017***	**0.013*****	0.018***	0.004***	**0.004*****	0.122***	0.247***	**0.173*****
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.016)	(0.016)	(0.017)
Av. School Rating	0.041***	0.024***	**0.017*****	0.015***	0.008***	**0.006*****	0.524***	0.316***	**0.231*****
	(0.0004)	(0.0004)	(0.0004)	(0.0002)	(0.0003)	(0.0003)	(0.005)	(0.006)	(0.007)
City Fixed Effects?	-	X	-	-	X	-	-	X	-
Neighborhood Fixed Effects?	-	-	X	-	-	X	-	-	X
Observations	54,654	54,654	54,654	54,654	54,654	54,654	54,654	54,654	54,654
R^2	0.702	0.872	0.910	0.763	0.862	0.881	0.339	0.581	0.659
Adjusted R^2	0.702	0.872	0.909	0.763	0.862	0.881	0.339	0.580	0.658

Note: *p**p***p<0.01

Table 7. Non-Linear Effects of School Quality

	Dependent variable:		
	log(price)	log(rent)	pr
	(1)	(2)	(3)
Av. School Rating	0.005**	-0.006***	0.237***
	(0.002)	(0.002)	(0.040)
Av. School Rating^2	0.001***	0.001***	-0.0004
	(0.0002)	(0.0001)	(0.003)
Neighborhood Fixed Effects?	X	X	X
Observations	54,654	54,654	54,654
R^2	0.910	0.881	0.659
Adjusted R^2	0.909	0.881	0.658
Residual Std. Error (df = 54469)	0.079	0.062	1.429
F Statistic (df = 184; 54469)	2,978.892***	2,201.941***	572.693***

Note: Housing attributes are hidden from regression output *p**p***p<0.01

Fig 3. Non-linear effects of Average School Quality

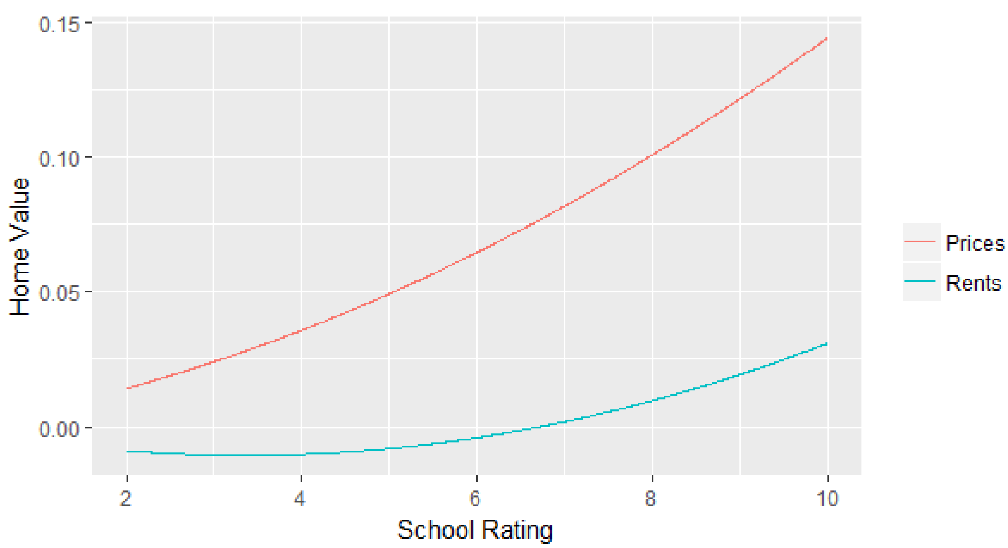


Table 8. Effects of Home Attributes on Prices, Segmented by Price Quartile

	Dependent variable:			
	log(Price)			
	(Q1)	(Q2)	(Q3)	(Q4)
# Bedrooms	0.009^{***}	-0.001^{**}	-0.001^{***}	-0.004^{***}
	(0.001)	(0.0005)	(0.001)	(0.001)
# Bathrooms	0.014^{***}	0.007^{***}	0.006^{***}	0.022^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Sqr. Ft. (100 ft.)	0.021^{***}	0.001^{**}	0.004^{***}	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Sqr. Ft.^{2} (100 ft.)	-0.0003^{***}	0.0001^{***}	0.00005^{***}	0.0003^{***}
	(0.00003)	(0.00002)	(0.00001)	(0.00001)
Lot Size (100 ft.)	0.005^{***}	0.001^{***}	0.0004^{***}	0.001^{***}
	(0.0001)	(0.00005)	(0.00005)	(0.0001)
Lot Size^2 (100 ft.)	-0.00002^{***}	-0.00000^{***}	0.00000^{**}	0.00000^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Age	-0.002^{***}	-0.0004^{***}	-0.001^{***}	-0.004^{***}
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Age^2	0.00002^{***}	0.00000^{***}	0.00001^{***}	0.00003^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
# Stories	-0.015^{***}	-0.001	-0.001	-0.007^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
#Parking Spaces	-0.0005	-0.0005	0.0004	0.0003
	(0.001)	(0.0004)	(0.0004)	(0.001)
Av. Distance to School	0.001	-0.001	0.009^{***}	0.019^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Av. School Rating	$\mathbf{0.011}^{***}$	$\mathbf{0.005}^{***}$	$\mathbf{0.007}^{***}$	$\mathbf{0.003}^{***}$
	(0.0005)	(0.0003)	(0.0004)	(0.001)
Neighborhood Fixed Effects?	X	X	X	X
Observations	13,554	13,644	13,595	13,689
R^2	0.926	0.976	0.972	0.936
Adjusted R^2	0.925	0.976	0.972	0.935
Note:				*p**p***p<0.01

Table 9. Effects of Home Attributes on Rents, Segmented by Price Quartile

	Dependent variable:			
	log(Rent)			
	(Q1)	(Q2)	(Q3)	(Q4)
# Bedrooms	0.027***	0.015***	0.011***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
# Bathrooms	0.027***	0.018***	0.014***	0.026***
	(0.001)	(0.001)	(0.001)	(0.001)
Sqr. Ft. (100 ft.)	0.040***	0.024***	0.024***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Sqr. Ft.^{2} (100 ft.)	-0.001***	-0.0003***	-0.0003***	0.0001***
	(0.00003)	(0.00002)	(0.00002)	(0.00001)
Lot Size (100 ft.)	0.002***	0.001***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Lot Size^2 (100 ft.)	-0.00001***	-0.00000***	-0.00000***	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Age	-0.0002	0.0003**	0.0002*	-0.001***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Age^2	-0.00000***	-0.00001***	-0.00001***	-0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
# Stories	-0.008***	-0.006***	-0.003***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
#Parking Spaces	0.00002	-0.001**	0.001	-0.0003
	(0.001)	(0.0005)	(0.0005)	(0.001)
Av. Distance to School	-0.001	0.0002	0.001	0.008***
	(0.001)	(0.001)	(0.001)	(0.002)
Av. School Rating	**0.004*****	**0.003*****	**0.005*****	**0.003*****
	(0.0004)	(0.0004)	(0.0004)	(0.001)
Neighborhood Fixed Effects?	X	X	X	X
Observations	13,554	13,644	13,595	13,689
R^2	0.868	0.906	0.906	0.869
Adjusted R^2	0.866	0.904	0.905	0.867
Note:				*p**p***p<0.01

Table 10. Effects of Home Attributes on Price-Rent Ratios, Segmented by Price Quartile

	Dependent variable:			
	Price-Rent Ratio			
	(Q1)	(Q2)	(Q3)	(Q4)
# Bedrooms	-0.330***	-0.321***	-0.263***	-0.088***
	(0.017)	(0.014)	(0.015)	(0.020)
# Bathrooms	-0.249***	-0.244***	-0.198***	-0.120***
	(0.026)	(0.023)	(0.023)	(0.028)
Sqr. Ft. (100 ft.)	-0.378***	-0.490***	-0.482***	-0.282***
	(0.020)	(0.017)	(0.016)	(0.016)
Sqr. Ft.^{2} (100 ft.)	0.009***	0.010***	0.009***	0.006***
	(0.001)	(0.0005)	(0.0004)	(0.0003)
Lot Size (100 ft.)	0.045***	-0.003**	-0.004***	0.004**
	(0.001)	(0.001)	(0.001)	(0.002)
Lot Size^2 (100 ft.)	-0.0001***	0.00004***	0.00004***	0.00003***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Age	-0.037***	-0.012***	-0.022***	-0.058***
	(0.004)	(0.003)	(0.003)	(0.003)
Age^2	0.0003***	0.0002***	0.0003***	0.001***
	(0.00003)	(0.00003)	(0.00003)	(0.00003)
# Stories	-0.111***	0.108***	0.038*	-0.074**
	(0.025)	(0.020)	(0.021)	(0.032)
#Parking Spaces	-0.008	0.009	-0.016	0.021
	(0.013)	(0.011)	(0.012)	(0.016)
Av. Distance to School	0.036	-0.026	0.151***	0.237***
	(0.024)	(0.023)	(0.026)	(0.036)
Av. School Rating	**0.133*** **	**0.041*** **	**0.031*** **	**-0.013**
	(0.010)	(0.009)	(0.011)	(0.020)
Neighborhood Fixed Effects?	X	X	X	X
Observations	13,554	13,644	13,595	13,689
R^2	0.736	0.814	0.808	0.660
Adjusted R^2	0.733	0.812	0.806	0.656
Note:				*p**p***p<0.01

Table 11. Effect of School Quality at the Grade Level, Regressed Using Fixed Effects to Control for Differences in School Assignment

	Dependent variable:								
	log(price)			log(rent)			pr		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Elementary School Rating	0.006^{***} (0.0003)			0.003^{***} (0.0002)			0.080^{***} (0.005)		
Middle School Rating		0.004^{***} (0.001)			0.002 (0.001)			0.047^{*} (0.028)	
High School Rating			0.002^{**} (0.001)			0.0005 (0.001)			0.033^{*} (0.018)
Neighborhood Fixed Effects?	X	X	X	X	X	X	X	X	X
Elementary School Fixed Effects?	-	X	X	-	X	X	-	X	X
Middle School Fixed Effects?	X	-	X	X	-	X	X	-	X
High School Fixed Effects?	X	X	-	X	X	-	X	X	-
Observations	54,654	54,654	54,654	54,654	54,654	54,654	54,654	54,654	54,654
R^2	0.921	0.941	0.941	0.887	0.898	0.898	0.683	0.720	0.720
Adjusted R^2	0.921	0.940	0.941	0.887	0.897	0.897	0.681	0.717	0.718

Note: Housing attributes are hidden from regression output *p**p***p<0.01

Table 12. Effect of School Quality at the Grade Level, Regressed Simultaneously

	Dependent variable:		
	log(Price)	log(Rent)	Price-Rent Ratio
	(1)	(2)	(3)
Elementary School Rating	0.006^{***}	0.002^{***}	0.071^{***}
	(0.0003)	(0.0002)	(0.005)
Middle School Rating	0.010^{***}	0.005^{***}	0.120^{***}
	(0.0004)	(0.0003)	(0.007)
High School Rating	0.001^{***}	-0.001^{***}	0.041^{***}
	(0.0004)	(0.0003)	(0.007)
Neighborhood Fixed Effects?	X	X	X
Observations	54,654	54,654	54,654
R^2	0.910	0.882	0.660
Adjusted R^2	0.910	0.881	0.658

Note: Housing attributes are hidden from regression output $^{*}p^{**}p^{***}p<0.01$

Table 13. Housing Attributes & Homeownership Rates

	Dependent variable:	
	Owner Occupied	
	Linear Probability Model	*Logit Model*
	(1)	(2)
# Bedrooms	0.002	0.015
	(0.003)	(0.021)
# Bathrooms	0.007^{*}	0.048
	(0.004)	(0.031)
Sqr. Ft. (per 100)	0.021^{***}	0.136^{***}
	(0.001)	(0.011)
Sqr. Ft.^{2} (100 ft.)	-0.0003^{***}	-0.002^{***}
	(0.00003)	(0.0002)
Lot Size (100 ft.)	0.001^{***}	0.008^{***}
	(0.0002)	(0.002)
Lot Size^2 (100 ft.)	-0.00000^{***}	-0.00003^{***}
	(0.00000)	(0.00001)
Age	0.003^{***}	0.022^{***}
	(0.0005)	(0.003)
Age^2	-0.00003^{***}	-0.0002^{***}
	(0.00000)	(0.00003)
# Stories	-0.012^{***}	-0.076^{***}
	(0.004)	(0.029)
#Parking Spaces	-0.003	-0.025
	(0.002)	(0.016)
Av. Distance to School	0.002	0.017
	(0.004)	(0.033)
Av. School Rating	0.001	0.007
	(0.002)	(0.013)
Price-Rent Ratio	-0.010^{***}	-0.055^{***}
	(0.001)	(0.008)
Observations	54,654	54,654
R^2	0.034	
Adjusted R^2	0.031	
Log Likelihood		-23,186.030
Akaike Inf. Crit.		46,742.070
Note:		*p**p***p<0.01

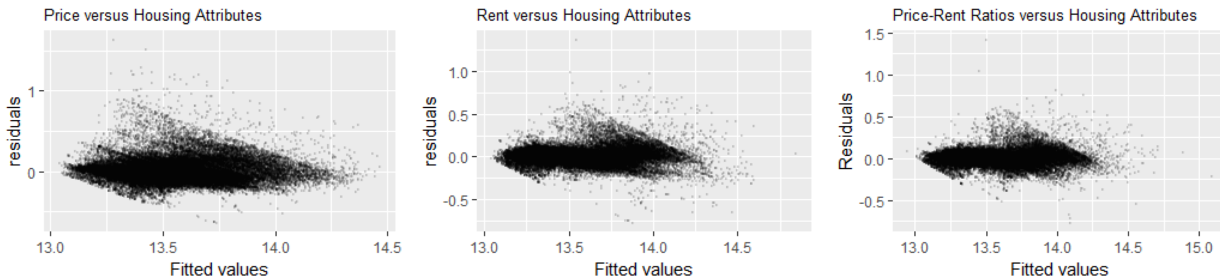
Table 14. Regression Results from Alternate Specifications

	Dependent variable:					
	Alt. Spec 1			*Alt. Spec. 2*		
	log(Price)	log(Rent)	Price-Rent Ratio	log(Price)	log(Rent)	Price-Rent Ratio
	(1)	(2)	(3)	(4)	(5)	(6)
Av. School Rating	0.018***	0.006***	0.254***	0.019***	0.007***	0.252***
	(0.001)	(0.001)	(0.020)	(0.002)	(0.002)	(0.038)
Neighborhood Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,580	7,580	7,580	3,352	3,352	3,352
R^2	0.902	0.883	0.638	0.884	0.822	0.574
Adjusted R^2	0.900	0.880	0.629	0.879	0.813	0.553

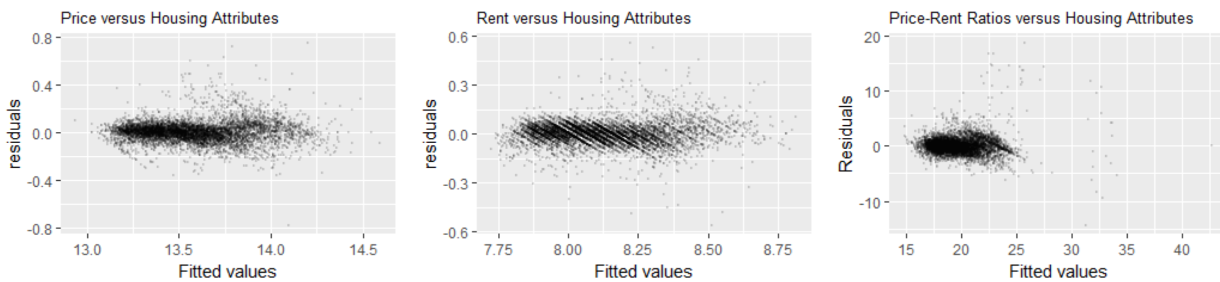
Note: Housing attributes are hidden from regression output *p**p***p<0.01

Figure 4. Residual Plots for Alternate and Full Data Specifications

Full Data



Alt. Spec. 1



Alt. Spec. 2

11. References

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, *64*(5), 416-424.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*(2), 577-599.

Fack, G., & Grenet, J. (2008). *Do Better Schools Raise Housing Prices? Evidence from Paris School Zoning, Paris School of Economics*. mimeo.

Davidoff, I., & Leigh, A. (2008). How much do public schools really cost? Estimating the relationship between house prices and school quality. *Economic Record*, *84*(265), 193-206.

Bayer, P., & McMillan, R. (2005). *Choice and competition in local education markets* (No. w11802). National Bureau of Economic Research.

Gibbons, S., & Machin, S. (2003). Valuing English primary schools. *Journal of urban economics*, *53*(2), 197-219.

Bogart, W. T., & Cromwell, B. A. (2000). How much is a neighborhood school worth?. *Journal of urban Economics*, *47*(2), 280-305.

Chiodo, A., Hernández-Murillo, R., & Owyang, M. T. (2010). Nonlinear effects of school quality on house prices. *Federal Reserve Bank of St. Louis Review*, *92*(May/June 2010).

Gibbons, S., Machin, S., & Silva, O. (2009). *Valuing School Quality Using Boundary Discontinuities, Centre for Economic Performance*. LSE, mimeo.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, *24*(3), 1141-1177.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis*, *19*(2), 141-164.

Brasington, D., & Haurin, D. R. (2006). Educational outcomes and house values: A test of the value added approach. *Journal of Regional Science*, *46*(2), 245-268.

Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, *18*(3), 253-273.

Himmelberg, C., Mayer, C., & Sinai, T. (2005). Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives*, *19*(4), 67-92.

Gallin, J. (2008). The long-run relationship between house prices and rents. *Real Estate Economics*, *36*(4), 635-658.

Kishor, N. K., & Morley, J. (2015). What factors drive the price–rent ratio for the housing market? A modified present-value analysis. *Journal of Economic Dynamics and Control*, *58*, 235-249.

Beracha, E., & Johnson, K. H. (2012). Lessons from over 30 years of buy versus rent decisions: is the American dream always wise?. *Real Estate Economics*, *40*(2), 217-247.

Campbell, S. D., Davis, M. A., Gallin, J., & Martin, R. F. (2009). What moves housing markets: A variance decomposition of the rent–price ratio. *Journal of Urban Economics*, *66*(2), 90-102.

Pancak, K. A. (2017). Variation in Local House Price–Rent Ratios. *Journal of Housing Research*, *26*(1), 79-94.

Lee, C., & Park, K. (2018). Analyzing the rent-to-price ratio for the housing market at the micro-spatial scale. *International Journal of Strategic Property Management*, *22*(3), 223-233.

Beracha, E., & Hardin III, W. G. (2018). The Capitalization of School Quality into Renter and Owner Housing. *Real Estate Economics*, *46*(1), 85-119.

Gruber, J. (2016). *Public finance and public policy*. Macmillan.

Hattapoglu, M., & Hoxha, I. (2014). The Dependency of rent-to-price ratio on appreciation expectations: An empirical approach. *The Journal of Real Estate Finance and Economics*, *49*(2), 185-204.

Bracke, P. (2015). House prices and rents: microevidence from a matched data set in Central London. *Real Estate Economics*, *43*(2), 403-431.

Chen, J., & Ni, S. (2010). Estimating Estate-Specific Price-to-Rent Ratios in Shanghai and Shenzhen: A Bayesian Approach. *Economics publications (MU).*

Bram, J. (2012). To buy or not to buy? The changing relationship between Manhattan rents and home prices. *Current issues in Economics and Finance*, *18*(9).

Chao Yue Tian. (2017). Are CAP Rates Constant within Housing Markets: Theory and Empirical Evidence. *Real Estate Finance (Aspen Publishers Inc.), 34*(2), 64–76.