12-11-2021

Non-linear and weakly monotonic relationship between school quality and house prices

Shishir Mathur
San Jose State University, mathurshishir1@gmail.com

Recommended Citation

Shishir Mathur. "Non-linear and weakly monotonic relationship between school quality and house prices" *Land Use Policy* (2021). https://doi.org/10.1016/j.landusepol.2021.105922

Land Use Policy 113 (2022) 105922

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

Non-linear and weakly monotonic relationship between school quality and house prices

Shishir Mathur

Urban and Regional Planning Department, San Jose State University, WSQ 216, One Washington Square, San Jose, CA 95192-0185, USA

ARTICLE INFO

Keywords:
Land use policy
School quality
Spatial hedonic regression
Housing
Urban economics

ABSTRACT

This study provides evidence for a non-linear and weakly monotonic relationship between school quality and house prices. Using Fremont, California, as the study area, the regression analysis shows that homeowners are unwilling to pay a premium for an increase in school quality from low to medium quality. However, they are willing to pay a) a large premium when all schools are top-quality schools and b) a premium for access to nationally-renowned schools, which is in addition to the premium for top-quality schools. These findings have important land use policy significance because they provide new insights into the homeowner's residential location choice and highlight the need to consider school quality in a jurisdiction's land use and zoning decisions.

1. Introduction

Many countries have adopted an attendance-area-based approach to school education. Under this approach, households need to send children to designated neighborhood public schools to avail free education. This approach leads to school quality becoming a major factor in households' location choice and, in turn, to the capitalization of school quality on house prices and rents. The capitalization occurs because households try to outbid each other to locate in areas with high-quality schools. Since higher-income households can outbid lower-income ones, this approach can lead to the concentration of the latter in areas with good schools. This concentration of higher-income households can create a positive feedback loop if higher housing prices lead to more revenues for the school district (for example, through increased tax revenues and parent donations), which, in turn, further improves the school quality; or, through the peer group effect where the wealthier kids generate better educational outcomes for their peers.[1] Such concentration is likely high for regions with very uneven school quality, a lack of housing with access to high-quality schools, and high household incomes. The San Francisco Bay Area of northern California in the US is one such region.

Such household sorting could lead to educational, economic, and social inequities when left to market forces because high house prices or rents restrict children from lower-income households from accessing high-quality education. Furthermore, empirical studies have shown strong linkages between the measures of education quality and earning potential (Hanushek, 2002; Card and Krueger, 1992, 1996); therefore, such a household sorting could worsen economic disparities between the rich and the poor. Finally, to the extent race/ethnicity and income might be highly correlated in a region, such sorting would likely concentrate minority communities in areas with low-quality schools (Mathur, 2017). These inequities-related concerns have spawned a large body of empirical studies. While US-focused initially, the more recent studies are from other parts of the world, mainly China, as regions in that country have started employing school-attendance-zone-based policy for admissions into public schools. For example, see Wen et al. (2017) study of Hangzhou, Hu et al. (2020) of Shanghai, and Zhang et al. (2020) meta-regression analysis of 38 China-focused studies.

Public action might be needed to address such inequities. Given the inequities' spatial nature and the fact that they are tied to the land, land-use-policy-related interventions might be required. Such interventions could include up-zoning in areas with high-quality schools so that more housing units can be provided in response to the demand for high-quality schools. Indeed, empirical studies show that inelastic housing supply increases school quality's house price premiums (Hilber and Mayer, 2009; Xiang et al., 2018; Zhang and Chen, 2018). Furthermore, targeted government subsidies might be needed to develop affordable housing for low-income households in areas with high-quality schools. Finally, land-use planners might need to address opposition from high-income property owners in such areas since these property owners are likely to resist up-zoning and the provision of affordable housing in their neighborhoods (He, 2017). However, the first step is to identify the school-quality-related house price premium.

The vast majority of empirical studies that estimate school quality's

[1] I thank an anonymous referee for suggesting this feedback loop.

https://doi.org/10.1016/j.landusepol.2021.105922
Received 2 January 2021; Received in revised form 1 December 2021; Accepted 2 December 2021
Available online 11 December 2021

impact on house prices assume a linear relationship between the two (Turnbull and Zheng, 2019). The few studies that account for non-linearity still assume a strong monotonic relationship, meaning that as school quality increases, so do the house prices (for example, see Chiodo et al., 2010; Cheshire and Sheppard, 2004). Please see the "Literature review" section for a fuller discussion.

I advance this line of research by using Fremont, California, as the study area to explore the possibility of a non-linear and weakly monotonic relationship between school quality and house prices. For example, such a relationship might exist if house prices or rents do not increase if school quality increases from low to medium quality.

The remainder of this paper is divided into four sections. In the next section, I review the literature. Focusing on empirical research, I identify the research gaps this study seeks to address. Next, in the "Research questions, study area, and data" section, I identify the specific research questions this study aims to answer, provide an overview of the study area, and describe the data. In the "Methods" section, I describe the methods used in this study and the robustness checks employed. In the "Results and major findings" section, I report the regression models' key findings, including the estimates of the impact of school quality on house prices. Finally, I conclude the paper by summarizing key findings and suggesting their potential land-use policy implications.

2. Literature review

2.1. Research design and model specification

Omitted variable (OV) bias is the primary concern that needs to be addressed while empirically estimating school quality's impact on house prices. That is, to demonstrate that the school quality, not other factors correlated with it, such as regional, jurisdictional, neighborhood-level, or policy-related influences, impacted house prices. Addressing endogeneity is the second concern. That is, to ensure that the school quality impacted house prices, not the other way round. The latter could occur, for example, if higher-income households move into a neighborhood for factors other than schools (for example, proximity to jobs or better quality neighborhoods), and then the school quality improves. In such a case, an increase in school quality is not the cause of house price increase. Indeed, an increase in house prices leads to better school quality.

A few studies have attempted to address endogeneity through econometric methods such as two-stage least squares (2SLS) and the instrumental variable (IV) regression techniques (for example, see Zahirovic-Herbert and Turnbull, 2008; Rosenthal, 2003; Bayer et al., 2007). Other studies have used novel research designs that address both endogeneity and OV bias problems. For example, Zahirovic-Herbert and Turnbull (2008) uses changes in school attendance zone boundaries. Wen et al. (2017) uses Hangzhou, China's enactment of school attendance zone policy to examine the policy's effect by including pre-, during, and post-policy implementation periods in their econometric analysis. Happily, a meta-analysis of 50 years' worth of empirical studies conducted by Turnbull and Zheng (2019) finds that efforts to address endogeneity do not influence the magnitude of school quality's house price impacts.

A large body of literature has devoted close attention to the OV bias problem and attempted to address it through research design and model specification, that is, to include all possible independent variables in a regression model that could impact house prices and are correlated with school quality. Apart from the research design used by Zahirovic-Herbert and Turnbull (2008) and Wen et al. (2017) discussed above; several studies use the boundary discontinuity design (BDD) approach pioneered by Black (1999). Under the BDD approach, only houses within a specific distance (for example, 0.25 mile) on either side of a school attendance zone boundary are included in the analysis, assuming that neighborhood-level and other locational attributes remain unchanged within this distance band. However, recent studies have critiqued this assumption and called for explicit control of neighborhood-level and locational attributes. Many such studies control neighborhood-level effects by using only Census data that capture the residents' socio-economic demographic characteristics. Such characteristics include the residents' race/ethnicity, education level, and age distribution (for example, see Chiodo et al., 2010; Zahirovic-Herbert and Turnbull, 2008). While other, more recent studies also include distance to various amenities and disamenities to control other locational attributes that impact house prices (see Hu et al., 2020; Kuroda, 2018). A nascent but growing body of literature mitigates OV bias by running spatial econometric models (see Mathur, 2017; Peng, 2019; Qui et al., 2020; Rajapaksa et al., 2020; Wen et al., 2019; 2018; 2017).

The data's spatial nature raises the likelihood of two types of spatial dependence: spatial error and spatial lag dependence. Under the former, the error terms may be correlated across space, thereby violating the assumption of uncorrelated error terms in OLS. This violation results in biased coefficient estimates and is often due to omitted spatial variables. For example, such biased estimates could be due to the omitted neighborhood-level variables. With spatial lag dependence, the dependent variable for an observation in one location could be affected by the dependent and independent variables for observations in other locations (Sedgley et al., 2008) because the sale price of a house might be influenced by the sale price and characteristics of houses sold in its vicinity. The presence of spatial lag dependence violates the assumptions of uncorrelated errors and the independence of observations, and it could lead to biased and inefficient estimates. Therefore, checking and addressing spatial dependence is necessary to mitigate the OV problem highlighted in the literature (if spatial error dependence is found) and the data's underlying spatial nature.

2.2. Relationship between school quality and house prices

2.2.1. Assumption of linearity and strong monotonic relationship between school quality and house prices

A meta-analysis of 56 US-focused studies conducted by Turnbull and Zheng (2019) finds that a large majority of research assumes a linear relationship between school quality and house prices. Chiodo et al. (2010) highlights the drawback of this assumption and notes that such an assumption "underestimates the premium at high levels of school quality and overestimates the premium at low levels of school quality" (Chiodo et al., 2010, 186). This study addresses non-linearity by including squared and cubed transformations of the school quality variable. Turnbull and Zheng (2019) note that apart from Chiodo et al. (2010), Cheshire and Sheppard (2004) is the only other study to address non-linearity, which it does by applying Box-Cox transformation to the hedonic regression model. However, both of the studies mentioned above still assume a strong monotonic relationship between school quality and house prices. For example, they do not account for a) the possibility that house prices or rents might not increase at all if school quality increases from low to medium quality, and b) sudden increases in house price premiums, for example, for top-quality or nationally-renowned schools.

In the broader literature on hedonic price modeling, non-parametric techniques have been advocated to address the larger issue of determining the functional form of a hedonic regression model since these techniques do not impose any functional form a priori. The menu of such techniques includes locally weighted regression, kernel density estimation, spline smoothing, series approximators, and nearest neighbors (see Anglin and Gencay, 1996; Coulson, 1992; Pace, 1993; Pace, 1995; Preez, Lee and Sale, 2013; McMillen and Redfearn, 2007). In addition, machine learning and data mining tools in this category include decision trees (for example, random forest regression) and support vector regression. While these techniques arguably lead to better model prediction, many of them require large data sets, are more suited for spatial data covering large geographies, and/or are unable to interpret the impact of specific independent variables on the dependent variable, for example, in the case of random forest regression (Waddell and Besharati-Zadeh, 2020).

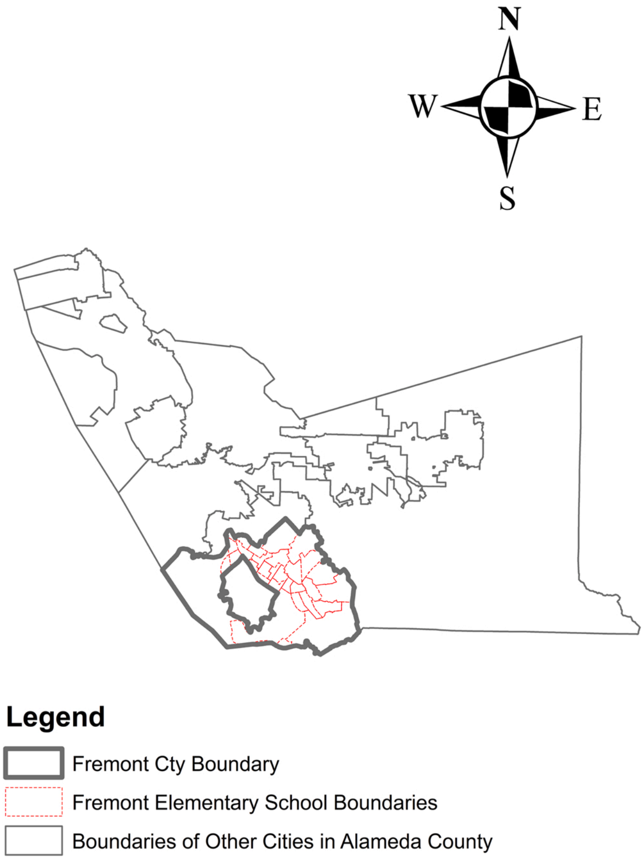

Fig. 1. Location map.

2.2.2. *Use of elementary school quality as an indicator of overall school quality or averaging elementary, middle, and high school quality*

Turnbull and Zheng's (2019) meta-analysis finds that a large majority of studies use the quality measures of elementary schools as proxies for overall school quality, thereby assuming that the middle and high schools are of the same quality as the elementary school. Studies that consider the quality of middle and high schools usually average the quality of each school. For example, Beracha and Hardin III (2018) averages the grades assigned by the Florida Department of Education to the elementary, middle, and high schools. Such averaging assumes that the impact of each school-level is the same on house prices. Recent studies have begun to parse the effect of the quality of various levels of schools on house prices. For example, Hu et al. (2020) creates one dummy variable each for elementary and middle school quality to assess the schools' impact on rental values in Shanghai, China. This study divides school quality into two categories: ordinary and high.

2.3. *Research gaps*

Review of the literature identifies three main research gaps. First, since most studies have only considered elementary school quality as a proxy for overall school quality, there is a need to parse the effect of various levels of schools—elementary, middle, and high—on house prices.

Second, even those studies that assess the house price impacts of various levels of schools assume that these schools impact house prices independent of each other. Research has not attempted to assess whether parents make trade-offs while choosing the bundle of elementary, middle, and high schools. For example, are parents more likely to choose a bundle comprising a high-quality elementary school and medium-quality middle and high school, or would they choose a bundle of low-quality elementary school and high-quality middle and high schools?

Third, almost all studies assume a linear or a strongly monotonic relationship between school quality and house prices. Hence, more research is needed to test this assumption. Finally, there is a debate in the literature on whether aggregate, school-district-level, measure of school quality is better compared to the measures that assess school quality of individual schools within that district. For example, Turnbull and Zheng (2019) notes that while Downes and Zabel (2002) argues that measures of quality of neighborhood schools are more effective in measuring the impact of school quality on house prices compared to measures of the quality of school district, Crone (2006) argues the opposite. By examining the effect of the bundle of neighborhood-level schools on house prices in one school district, my research informs this on-going debate.

3. Research questions, study area, and data

3.1. *Research questions*

This study seeks to fill the research gaps identified in the "Literature review" section. Specifically, it seeks to answer the following research questions:

a) Controlling for other factors, does the school quality and house prices have a weak monotonic relationship?

b) Is the school quality's house price premium influenced by the interaction between the quality of different schools (elementary, middle, and high)?

3.2. *Study area*

Fremont, CA, is the study area for this research. The city is situated in the Alameda County of the San Francisco Bay Area region of northern California. See Fig. 1 for the location of Fremont within Alameda County. With the year 2010 population of approximately 241,000, Fremont is an Asian-majority (about 59%), largely single-family, suburban city with a homeownership rate of roughly 61% (U.S. Census Bureau, 2019). The total number of school-going children (I use the age group 5–17 for this purpose) increased significantly between 2010 and 2019 in Fremont—from 37,955 to 55,214 (U.S. Census Bureau, 2019; Bay Area Census, n.d.). These numbers show that while most of the children in the 5–17 age group in the year 2010 would have crossed this age group by 2019, their numbers were more than compensated by "new" children. Therefore, households with school-going children were likely a large proportion of bidders for homes in Fremont during the study period (2012–2014).

Fremont is an ideal study area because one school district—Fremont Unified School District (FUSD)—serves the entire city, so the econometric models do not have to parse the jurisdiction-level effects from the school-quality effects. Furthermore, the FUSD has schools of varying quality allowing for fine-grained estimation of school quality's capitalization on house prices.

3.3. *Data description*

The dataset includes the sale price and the structural and locational attributes of 801 single-family houses sold during the period April 2012–March 2014.[2] The data for structural attributes include the number of bedrooms, number of bathrooms, size of the living space, lot size, most recent sale date, and the construction year. I used the sale date

[2] The study dataset includes single-family houses that can be owned as well as rented. I considered the option of including townhomes, condominiums, and apartments. However, the property characteristic data for these housing types were either of uneven quality or not available. Hence, I focused on single-family houses.

Table 1
Decile ranks of schools.

Elementary, Middle, and High School Combination	Decile Ranks	Sum of Decile Ranks
Blacow Elementary, Walters Middle, and Kennedy High (BWK)	5 6 4	15
Brier Elementary, Walters Middle, and Kennedy High (BRWK)	7 6 4	17
Durham Elementary, Walters Middle, and Kennedy High (DWK)	7 6 4	17
Grimmer Elementary, Horner Middle, and Irvington High (GMHI)	5 10 10	25
Green Elementary, Horner Middle, and Irvington High (GRHI)	7 10 10	27
Hirsch Elementary, Horner Middle, and Irvington High (HHI)	9 10 10	29
Weibel Elementary, Horner Middle, and Irvington High (WHI)	10 10 10	30
Glenmoor Elementary, Centerville Middle, and Washington High (GCW)	7 7 6	20
Maloney Elementary, Centerville Middle, and Washington High (MCW)	6 7 6	19
Parkmont Elementary, Centerville Middle, and Washington High (PCW)	10 7 6	23
Mission Valley Elementary, Hopkins Middle, and Mission San Jose High (MHM)	10 10 10	30
Oliveira Elementary, Thornton Middle, and American High (OTA)	7 8 8	21

data to calculate dummy variables representing the quarter of the sale year.

I gathered the data for various school-level attributes such as the students' race/ethnicity, average class sizes, student-teacher ratio, and percent of students provided with free or reduced-price lunch (an indicator of the students' family income); and on Academic Performance Index (API) score of all the schools. API is an output-based measure of school quality. The California State Department of Education calculated and reported API for all public schools in the state till 2014. It represented the results of testing over the entire course of a student's education—from grade 3–11. Therefore, API scores were a comprehensive measure of school quality. Till 2014, which includes the entire study period, these scores were prominently highlighted in real estate listings for homes in the areas with high API scores. Therefore, API scores were a well-known measure of school quality to both home buyers and sellers. The FUSD contains schools that were state-average in API scores to the nationally-renowned schools in the city's Mission San Jose area—a wide range of school quality. I obtained the attendance zone boundaries of all elementary, middle, and high schools from an on-line vendor that hosts the geographic information system (GIS) data for the FUSD.[3] These boundaries were then digitized using ArcGIS software and appended to each house parcel along with the data on school quality and school and locational attributes.

Blacow Elementary, Walter Middle, and Kennedy High schools comprise the lowest-quality bundle of elementary, middle, and high schools in the study area. These schools are in the fifth, sixth, and fourth deciles of all California public schools in quality, respectively, when ranked by API (California Department of Education, 2015). The nationally-renowned Mission Valley Elementary, Hopkins Middle, and Mission San Jose High schools are the highest quality schools in the study area. These top-quality schools are in the top, tenth, decile (i.e., the top 10%) of all California public schools when ranked by API. Table 1 provides the decile ranks for all the bundles of all elementary, middle, and high schools in the study area. The first number in the "Decile Ranks" column represents the decile rank of the elementary school, the second number of the middle school, and the third number of the high school. These school-quality bundles, operationalized through dummy variables, are used to measure school quality. To address the omitted variable (OV) bias, that is, to ensure that these dummy variables do not pick up the neighborhood and other locational effects, I gathered data for locational attributes. These locational attribute variables include those measuring the proximity of each house to various amenities and disamenities such as industrial and commercial uses, mobile homes, offices, and open spaces/parks; and b) US Census data at the block-level for various neighborhood-level attributes such as demographic and economic characteristic that include race/ethnicity, income, and percent renter population.

Table 2
Descriptive statistics for the continuous variables used in the final models. N = 801.

	Minimum	Maximum	Mean	Std. Deviation
Consumer price index-adjusted sale price of the house ($)	$90,621	$16,57,241	$6,67,690	$2,73,408
House size (square feet)	796	4102	1611	551
Lot size (square feet)	2910	25,370	7015	2571
Number of bedrooms	2.0	6.0	3.4	0.7
Number of bathrooms	1.0	4.5	2.1	0.6
Age of the house (year effective)	0.0	97.0	46.0	12.5
Percent renter population in the census block	0.0	100.0	23.8	17.2
Percent Asian population in the census block	0.0	100.0	41.9	19.4
Distance to nearest multi-family houses (in feet)	23	4005	797	768
Distance to nearest mobile homes (in feet)	42	16,485	7618	3705
Distance to nearest commercial use (in feet)	27	5753	1272	806
Distance to nearest industrial use (in feet)	23	5809	1400	833
Distance to nearest open space, including parks (in feet)	28	1947	523	367
Distance to nearest institutional use (in feet)	25	3758	933	598
Distance to nearest office use (in feet)	35	6065	1957	1160

I took two more steps to reduce the OV bias. First, to ensure that the school quality effects do not comingle with jurisdiction-level impacts (Bayer et al., 2007), I excluded the school attendance zones that share the city boundaries from the analysis. Therefore, the dataset includes data only for the houses located in school attendance zones that are entirely within the city boundary. Second, I included houses only within the 0.375 miles of either side of the elementary school boundaries.

The final dataset comprises 801 observations spread across 12 elementary schools and five middle and high schools. Table 2 provides descriptive statistics for the continuous variables used in the final model. The mean house price is $667,690, with a standard deviation of $273,408. I used the non-housing consumer price index (CPI) for the region to adjust the house price for inflation.

4. Methods

This study employs the hedonic regression approach (ordinary least squares [OLS] and spatial regression) to estimate the implicit price associated with the quality, Q, of the bundle of elementary, middle, and high schools (Q_{emh}). Therefore, the main estimation equation regresses the sale price of a single-family house i (SP_i) on its structural (STi) and locational attributes (L) in neighborhood j (L_j); and the quality of schools, Q_{emh}. ξ_i is the error term, which is assumed to be independent of

[3] In the US, public schools are divided into elementary, middle, and high schools. Elementary school usually consists of Kindergarten through fifth grade, middle school sixth to eighth grade, and high school ninth to twelfth grade.

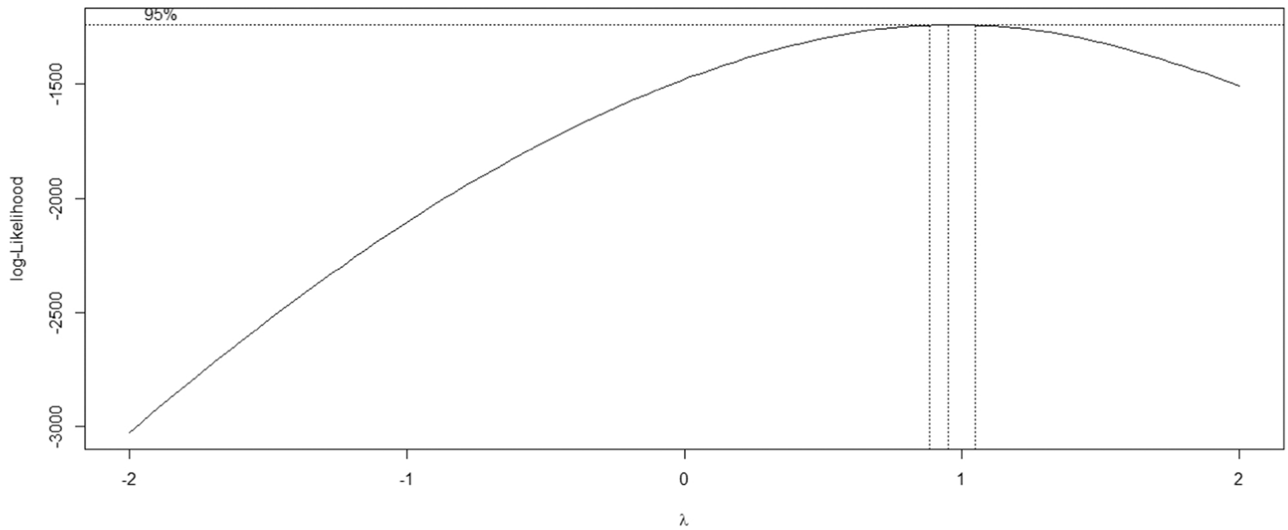

Fig. 2. Box-Cox transformation: Lambda Value.

Q_{emh}.

$$SP_i = \alpha_0 + ST_i + L_j + Q_{emh} + \xi_i \quad (1)$$

Estimating equation (1) using OLS regression assumes homoscedasticity, or constant variance of the error term, as shown in equation (2).

$$V(\xi_i) = \sigma^2 \text{ for all i.} \quad (2)$$

Violating this assumption could lead to biased standard errors of the coefficients, that is, an over- or under-estimation of the standard errors. Such violations typically occur when the error term's variance is a function of a vector of explanatory variables z_{ij} (see Eq. (3)). Indeed, the Breusch-Pagan test for heteroscedasticity indicates a non-constant variance of the error term for the preliminary OLS regression models estimated in this study. Therefore, I estimated the regression models with robust standard errors clustered at the school attendance zone level because the independent variable of interest, school quality, varies at this level.

$$\sigma_i^2 = \sigma^2 f\left(\alpha_0 + \sum_{j=1}^{P} \alpha_j z_{ij}\right) \quad (3)$$

Next, I conducted a Box-Cox transformation to identify the model's suitable functional form. The λ (lambda) value is very close to 1, indicating that the linear model specification is appropriate. See Fig. 2.

Further, OLS assumes the independence of explanatory variables. Specifically, the error terms are assumed not to correlate with each other. The temporal nature of the data in this study (spread over eight quarters—Spring 2012 to Winter 2014) increases the likelihood of temporal autocorrelation, the presence of which could lead to biased standard errors and thus reduce a model's explanatory power. Therefore, I conducted the Breusch–Godfrey test for serial correlation. This test's result was statistically insignificant, indicating an absence of serial correlation. Moreover, to ensure that multicollinearity did not affect the statistical significance or the coefficient of the school quality variables, I only included independent variables with a variance inflation factor (VIF) of less than 10.

After that, I checked for spatial dependence. The first step was to create a spatial weights matrix, W, to weight the sale price by accounting for the sales transactions' spatial and temporal proximity. Using the methodology employed by Di et al. (2010), I included the four sale transactions nearest to a given house to calculate the spatial weights. We further weighted the transactions by the proximity of the sale year. I gave a weight of 1 for transactions in the same year, a weight of 0.5 for the transactions two years apart, and a weight of 0.33 for the transactions three years apart. Finally, I row-standardized the spatial weights.

Table 3
Lagrange multiplier (LM) test results.

Test	p-value
LM test for error dependence (LMerr)	0.000001354
Simple LM test for a missing spatially lagged dependent variable (LMlag)	0.00003122
Test for error dependence in the presence of a missing lagged dependent variable (RLMerr)	0.000571
Test for a missing lagged dependent variable in the presence of error dependence (RLMlag)	0.01545

Second, I conducted the Moran's I test for spatial autocorrelation in residuals. The Moran's I value was statistically significant ($p = 9.05 \times 10^{-8}$). Next, I employed the Lagrange multiplier (LM) tests to ascertain the type of spatial dependence that the models exhibit: spatial lag, spatial error, or both (Anselin, 1988). I used the following LM tests: the simple LM test for error dependence (LMerr) and for a missing spatially lagged dependent variable (LMlag); I used the RLMerr test for error dependence in the presence of a missing lagged dependent variable and the RLMlag test for a missing lagged dependent variable in the presence of error dependence (Bivand and Bernat, 2011).

The LM tests indicated both spatial lag and error dependence (see the low p-values for RLMerr and RLMlag in Table 3). Therefore, we ran both spatial error and spatial lag regression models and selected the model with the higher log likelihood to report the regression results, which in this case was the spatial error model. The spatial error model equation was estimated as follows:

$$\begin{aligned} P_i &= \alpha_0 + ST_i + L_j + Q_{emh} + \xi \\ \xi &= \lambda W \xi + \varepsilon \end{aligned} \quad (4)$$

where ξ is a vector of error terms that is spatially weighted by using the weights matrix, W;.

λ is an autoregressive parameter; and.

ε is a vector of uncorrelated error terms.

The spatial lag equation was estimated as follows:

$$P_i = \alpha_0 + \rho W P_i + ST_i + L_j + Q_{emh} + \xi \quad (5)$$

where.

WP is a spatially lagged dependent variable for the weights matrix, W, and.

ρ is a spatial autoregressive parameter.

Table 4
Regression results.

	Spatial Error Model			OLS Model		
Variables	Coefficient	Sig.	Std. Error	Coefficient	Sig.	Std. Error (clustered-robust standared error)
Structural Attributes						
House size (square feet)	169	***	13	171	***	22
Lot size (square feet)	11.0	***	2.0	10.7	***	1.8
Number of bedrooms	-5390		7344	-8117		7402
Number of bathrooms	10,710		9570	12,610		8540
Age of the house (year effective)	-3482	***	428	-3430	***	615
Locational Attributes						
Percent Asian population in the census block	86		319	23		336
Percent renter population in the census block	-195		247	-257		266
Distance to nearest multi-family houses (in feet)	16	*	9	15		14
Distance to nearest commercial use (in feet)	-1		8	-2		8
Distance to nearest industrial use (in feet)	18	***	7	20	**	10
Distance to nearest institutional use (in feet)	-13	*	7	-13		10
Distance to nearest office use (in feet)	8		5	9	**	4
Distance to nearest open space, including parks (in feet)	-9		11	-10		16
DisMobHome	2		3	3		7
School Dummies						
BRWK	1579		3056	2635		21,478
DWK	-6926		20,653	-5538		18,056
GMHI	29,692		24,192	32,021		32,248
GRHI	74,718	***	24,115	83,467	**	36,422
HHI	1,61,390	***	18,842	1,63,200	***	15,336
WHI	2,80,290	***	29,624	2,80,190	***	52,941
GCW	55,360	*	28,761	45,210		54,074
MCW	47,572		34,476	51,195		45,056
PCW	65,772	***	18,184	70,037	***	12,352
MHM	3,60,680	***	19,087	3,66,060	***	18,504
OTA	15,815		22,836	17,862		38,709
Quarter of Sale						
Summer 2012	17,753		12,956	17,825		13,345
Fall 2012	35,644	***	13,358	34,983	***	6856
Winter 2013	88,237	***	15,183	87,179	***	17,899
Spring 2013	1,51,450	***	13,320	1,49,840	***	12,894
Fall 2013	1,63,840	***	13,554	1,63,480	***	10,567
Summer 2013	1,80,370	***	12,758	1,76,010	***	10,056
Winter 2014	2,29,510	***	18,196	2,29,590	***	13,835
λ	0.25533	***				
Adjusted R^2				0.855		
N	801			801		

*** significant at $p = 0.01$ level; ** significant at $p = 0.05$ level; * significant at $p = 0.1$ level.

In the next section, I report the results of the spatial error model. The Table 4 reports results of both the spatial error model and the OLS model with the standard errors clustered at the school attendance zone level.

5. Results and major findings

5.1. Results

The adjusted R-square for the OLS model is 0.855, indicating that the model has a high degree of explanatory power. Below, I report the results of the spatial error model since the OLS model displays spatial error dependence.

The coefficients of all the variables significant at $p = 0.05$ level have expected signs. The increase in the size of the house and the lot increases house prices, while an increase in age decreases house prices. All but one coefficient of the quarter dummies, which capture the strength of the real estate market, are statistically significant; and the magnitude of the coefficients reflect the increasing strength of the housing market during the study period as it recovered from the 2008 recession. Among the variables measuring proximity to various amenities and disamenities, only the distance to industrial uses was statistically significant at $p = 0.05$ level. The house prices decreased with proximity to industrial uses.

The variables measuring neighborhood quality (percent Asians and percent renters) are statistically insignificant. Several of the school quality dummy variables are statistically significant, however. They all represent the bundles of high-quality elementary, middle, and high schools. The lowest quality school attendance zone for this study area was the referent category. The students living in this zone attend Blacow Elementary, Walters Middle, and Kennedy High schools, which have decile ranks of 5, 6, and 4, respectively. See Table 4 for the detailed results. An in-depth discussion of the school quality variables follows in the "Major findings" section.

To test whether endogeneity biases the house price impacts of school quality, I estimated a two-stage least square (2SLS) regression. I used the decile ranks of schools to create a new continuous variable capturing school quality called "Sum of Decile Ranks" which essentially adds the decile rank of each school that comprises the elementary-middle-high school bundle. For example, the sum of decile ranks of the bundle of Durham Elementary (decile rank 7), Walters Middle (decile rank 6), and Kennedy High (decile rank 4) equals 17 (7 +6 +4). See Table 1 for the sum of decile ranks for all the school bundles.

Since we expect SDECILE to be endogenous with house prices, in the first stage of 2SLS, I regressed SDECILE on a set of exogenous variables. These variables included percent renters, percent Asians, the student-teacher ratio for high schools and elementary schools lagged by one year, the percent of students receiving free or reduced-price meals in high schools and elementary schools, and the percent of Asian students in high schools and elementary schools lagged by one-year. In the second stage, I replaced SDECILE with the predicted values obtained from the first stage (see the Apihat variable in Table 5).

Next, I ran an OLS model with the same set of regressors as in 2SLS

Table 5
Regression results: 2SLS and OLS for checking endogeneity.

	2SLS			OLS Model		
Variables	Coefficient	Sig.	Std. Error	Coefficient	Sig.	Std. Error
Structural Attributes						
Apihat	20,200	***	1114			
SDECILE				16,140	***	1122
House size (square feet)	174	***	15	172	***	16
Lot size (square feet)	11.6	***	1.9	12.0	***	2.0
Number of bedrooms	-6528		8256	-9810		8754
Number of bathrooms	7659.0		10,560	13,720		11,180
Age of the house (year effective)	-3065	***	471	-2990	***	500
Locational Attributes						
Percent Asian population in the census block	1024	**	329	1776	***	340
Percent renter population in the census block	-41		271	-506	**	284
Distance to nearest multi-family houses (in feet)	37	***	7	24	**	7
Distance to nearest commercial use (in feet)	-3		8	-12		9
Distance to nearest industrial use (in feet)	36	***	7	44	***	7
Distance to nearest institutional use (in feet)	-20	**	7	-26	**	8
Distance to nearest office use (in feet)	11	*	5	16	**	5
Distance to nearest open space, including parks (in feet)	18		11	22	*	12
DisMobHome	6	***	1	8	***	1
Quarter of Sale						
Summer 2012	21,990		14,300	22,110		15,160
Fall 2012	32,430	*	14,560	33,440	*	15,430
Winter 2013	91,940	***	16,620	89,240	***	17,620
Spring 2013	1,58,100	***	14,740	1,60,200	***	15,620
Fall 2013	1,61,200	***	14,870	1,62,400	***	15,770
Summer 2013	1,61,500	***	14,170	1,61,800	***	15,020
Winter 2014	2,20,900	***	20,340	2,19,900	***	21,560
Adjusted R^2	0.83			0.81		
N	801			801		

*** significant at p = 0.01; ** significant at p = 0.05 level; * significant at p = 0.1 level.

and compared the 95% confidence interval (C.I.) for the "Sum of Decile Ranks" in the OLS model with the 95% C.I. for the Apihat variable from the 2SLS model and found that the intervals overlap. This overlap signifies that endogeneity is not a problem in this study's model. Specifically, the 95% C.I. for the Apihat in the 2SLS and the SDECILE in the OLS models is $18,016-$22,383 and $13,941-$18,339, respectively. See Table 5 for the regression results.

Table 6
Regression results: robustness check.

Variables	Coefficient	Sig.	Std. Error
Structural Attributes			
House size (square feet)	140	***	30
Lot size (square feet)	15.0	**	4.7
Number of bedrooms	-10,847		7010
Number of bathrooms	12,600		10,140
Age of the house (year effective)	2824	***	414
Locational Attributes			
Percent Asian population in the census block	-195		241
Percent renter population in the census block	-199		270
Distance to nearest multi-family houses (in feet)	-4		15
Distance to nearest commercial use (in feet)	-7		31
Distance to nearest industrial use (in feet)	19		19
Distance to nearest institutional use (in feet)	-28	**	12
Distance to nearest office use (in feet)	1		4
Distance to nearest open space, including parks (in feet)	-28		18
DisMobHome	-4		9
School Dummies			
BRWK	27,857		21,972
DWK	-14,065		16,795
GCW	82,939		64,707
MCW	78,743		50,508
OTA	40,187		44,963
GMHI	-19,798		37,236
Quarter of Sale			
Summer 2012	22,508		20,045
Fall 2012	42,117	***	7593
Winter 2013	79,207	***	17,904
Spring 2013	1,33,970	***	19,394
Fall 2013	1,69,460	***	12,328
Summer 2013	1,72,270	***	12,902
Winter 2014	2,23,400	***	21,299

Adjusted R-square = 0.632.
N = 394.
*** significant at p = 0.01 level; ** significant at p = 0.05 level; * significant at p = 0.1 level.

5.2. Major findings

The regression results, specifically, the school quality variables, provide the following major findings:

5.2.1. No price premium for an increase in school quality from low to medium

Compared to houses in the referent category [lowest school quality attendance zone—Blacow Elementary (decile rank 5), Walters Middle (decile rank 6), and Kennedy High school (decile rank 4)], households are not willing to pay more for low to moderate increases in school quality. We can see this lack of willingness in statistically insignificant coefficients (at p = 0.05 level) for BRWK (7, 6, and 4 decile ranks), DWK (7, 6, and 4 decile ranks), MCW (6, 7, and 6 decile ranks), GCW (7, 7, and 6 decile ranks), OTA (7, 8, and 8 decile ranks), and GMHI (5, 10, and 10 decile ranks) school attendance zones.

To further check the findings' robustness, I estimated another regression model that included data from only BWK (referent category), BRWK, DWK, GCW, GMHI, MCW, and OTA school attendance zones—all schools with low to medium quality. The model results showed that the coefficients for all the school dummies were statistically insignificant at p = 0.05 level, further reinforcing that homebuyers consider the combinations of elementary, middle, and high schools represented by these dummy variables similar in quality. See Table 6 for regression results.

5.2.2. Importance of elementary school quality

The statistically insignificant coefficient for the GMHI attendance zone (5, 10, and 10 decile ranks) indicates households' lack of willingness to pay a price premium when elementary school quality is low, even though the middle and high schools are high-quality. On the other hand,

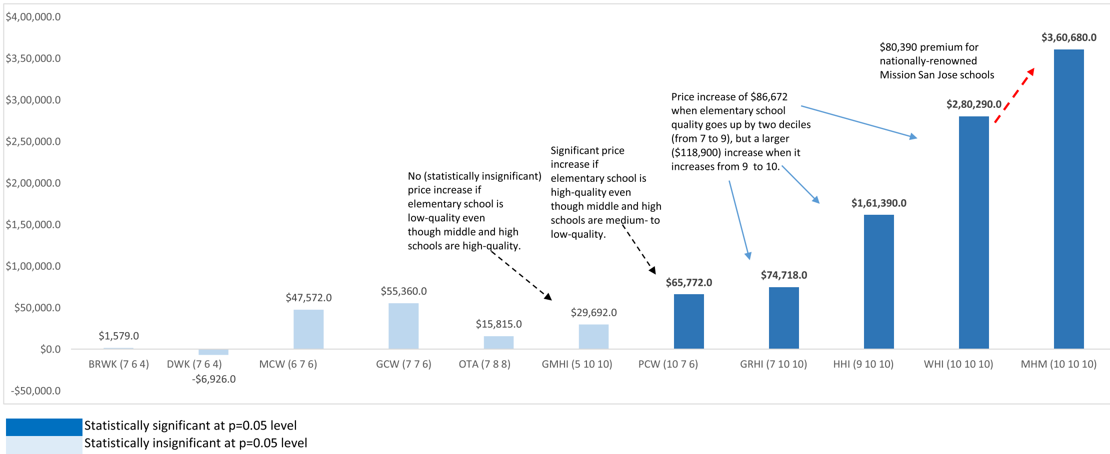

Fig. 3. House price premium by school quality.

the statistically significant positive coefficient for the PCW attendance zone (10, 7, and 6 decile ranks) indicates a school-quality price premium when elementary school is high-quality even though the middle and high schools are not. This emphasis on elementary schools could be because many first-time homebuyers with very young children might be buying houses in the study area. These buyers are willing to bid high for houses with high-quality elementary schools even if the middle and high schools are low- to medium-quality. Perhaps they hope that the middle and high schools' quality would improve by the time their children finish elementary school, or they plan to sell their houses at that time. Caetano (2019) provides another explanation by noting that the focus on elementary schools might be because parents of elementary-school-going children who face budget constraints might focus on the short-term by purchasing houses in areas with high-quality elementary schools.

Elementary schools' importance is further emphasized by the $86,672 difference in the GRHI and HHI attendance zones' coefficients ($161,390 coefficient of HHI minus $74,718 coefficient of GRHI)—a 13% price difference for this study's dataset. Both the attendance zones have high-quality (decile 10) middle and high schools. However, GRHI has a medium-quality elementary school (decile 7), while HHI has a high-quality elementary school (decile 9).

5.2.3. *Households are willing to pay a significant premium when all the three schools are top-quality*

The difference in the HHI and WHI's coefficients indicates that households are willing to pay a significant premium when all three schools are top-quality. The elementary school's decile rank increases only one decile—from 9 for HHI to 10 for WHI (from a high-quality to a top-quality school), and the associated price increase is $118,900 ($280,290 coefficient of WHI minus $161,390 coefficient of HHI)—an 18% increase for this study's dataset. This price increase is more than the $86,672 increase associated with a two-decile jump—from a medium-quality school of 7 decile rank in the case of GRHI to a high-quality school of 9 decile rank in the case of HHI—reflecting the large premium households are willing to pay for top-quality schools.

5.2.4. *A significant price premium is associated with nationally-renowned schools*

Mission San Jose schools, represented in this study by the MHM attendance zone, are nationally-renowned. The difference of $80,390 in the coefficients of WHI and MHM ($360,680 coefficient of MHM minus $280,290 coefficient of WHI)—both attendance zones of top-quality schools—indicates that households are willing to pay an additional significant premium of approximately $80,000 (12% of the average-priced house in this dataset) for nationally-renowned schools. See Fig. 3 for a graphical representation of the house premium of school quality.

6. Conclusions and policy implications

This study adds to the body of empirical studies estimating the impact of school quality on house prices by parsing the effect of the bundle of school quality on house prices along the school quality spectrum. It shows that homeowners might not be willing to pay a premium for an increase in school quality from low- to medium-quality. However, they are willing to pay a large premium for high-quality elementary schools when middle and high schools are high-quality too. Specifically, in this study's dataset, the house prices increase 13% when elementary school quality increases from average to high quality, and another 18% when it increases from high- to top-quality. Furthermore, homeowners are willing to pay a premium (12% in this study) to access nationally-renowned schools above and beyond the premium for top-quality schools.

The above findings have important land use policy implications since they provide new insights into the homeowners' residential location choice and highlight the need to consider school quality in a jurisdiction's land use and zoning decisions. Specifically, jurisdictions that aim to provide access to high-quality education to their residents need to make concerted efforts to zone for higher densities in areas with high-quality schools. However, since areas with high-quality schools are also likely to have high land prices, land-use planners need to proactively work with housing planners and policymakers to facilitate affordable housing in these areas. Finally, the existing residents, especially homeowners, might oppose up-zoning or the provision of affordable housing since these actions will likely dilute the school quality's house price premium which they paid at the time of buying houses (He, 2017)). Hence land-use planners and policymakers need to anticipate and address such resident opposition proactively.

References

Anglin, P., Gencay, R., 1996. Semiparametric estimation of hedonic price functions. J. Appl. Econ. 11, 633–648.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht, the Netherlands.

Bay Area Census, n.d. City of Fremont: Alameda County. ⟨http://www.bayareacensus.ca.gov/cities/Fremont.htm⟩ (Accessed 01 December 2021).

Bayer, P., Fernando, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. J. Polit. Econ. 115, 588–638.

Beracha, E., Hardin III, W., 2018. The capitalization of school quality into renter and owner housing. Real Estate Econ. 46, 85–119.

Bivand, R., Bernat, A., 2011. Lagrange multiplier diagnostics for spatial dependence in linear models. ⟨http://rss.acs.unt.edu/Rdoc/library/spdep/html/lm.LMtests.html⟩ (Accessed 3 November 2011).

Black, S., 1999. Do better schools matter? Parental valuation of elementary education. Q. J. Econ. 114, 577–599.

Caetano, G., 2019. Neighborhood sorting and the value of public school quality. J. Urban Econ. 114, 103193.

California Department of Education, 2015. State decile ranks-Highest Cut Scores. ⟨http://www.cde.ca.gov/ta/ac/ap/highestapiscore14.asp⟩ (Accessed 6 November 2015).

Card, D., Krueger, A., 1992. School quality and Black-White relative earnings: a direct assessment. Q. J. Econ. 107, 151–200.

Card, D., Krueger, A., 1996. Labor market effects of school quality: Theory and evidence. NBER Working Paper 5450. ⟨http://www.nber.org/papers/w5450.pdf⟩ (Accessed 2 November 2019).

Cheshire, P., Sheppard, S., 2004. Capitalising the value of free Schools: the impact of supply characteristics and uncertainty. Econ. J. 114, F397–F424.

Chiodo, A., Hernández-Murillo, R., Owyang, M., 2010. Nonlinear effects of school quality on house prices. Fed. Reserve Bank St. Louis Rev. 92, 185–204.

Coulson, E., 1992. Semiparametric estimates of the marginal price of floorspace. J. Real Estate Financ. Econ. 5 (1), 73–83.

Crone, T., 2006. Capitalization of the quality of local public schools: What do home buyers value? Working Paper, Federal Reserve Bank of Philadelphia. ⟨https://www.philadelphiafed.org/-/media/frbp/assets/working-papers/2006/wp06–15.pdf⟩ (Accessed 21 December 2020).

Di, W., Ma, J., Murdoch, J., 2010. An analysis of the neighborhood impacts of a mortgage assistance program: a spatial hedonic model. J. Policy Anal. Manag. 29, 682–697.

Downes, T., Zabel, J., 2002. The impact of school characteristics on house prices: Chicago 1987-1991. J. Urban Econ. 52, 1–25.

Hanushek, E., 2002. The long run importance of school quality. NBER Working Paper 9071. ⟨http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202002%20⟩ OurSchOurFuture.pdf (Accessed 11 November 2015).

He, S., 2017. A hierarchical estimation of school quality capitalisation in house prices in Orange County, California. Urban Stud. 54, 3337–3359.

Hilber, C.A.L., Mayer, C., 2009. Why do households without children support local public schools? Linking house price capitalization to school spending. J. Urban Econ. 65, 74–90.

Hu, L., He, S., Luo, Y., Su, S., Xin, J., Weng, M., 2020. A social-media-based approach to assessing the effectiveness of equitable housing policy in mitigating education accessibility induced social inequalities in Shanghai, China. Land Use Policy 94, 1–17.

Kuroda, Y., 2018. The effect of school quality on housing rents: evidence from Matsue city in Japan. J. Jpn. Int. Econ. 15, 16–25.

Mathur, S., 2017. The myth of "free" public education: impact of school quality on house prices in the Fremont Unified School District, California. J. Plan. Educ. Res. 37, 176–194.

McMillen, P., Redfearn C., 2007. Estimation, interpretation, and hypothesis testing for nonparametric hedonic house price functions. Working Paper. Lusk Center for Real Estate, University of Southern California. ⟨https://lusk.usc.edu/sites/default/files/working_papers/wp_2007–1007.pdf⟩ (Accessed 01 October 2021).

Pace, R., 1993. Nonparametric methods with applications to hedonic models. J. Real. Estate Financ. Econ. 7, 185–204.

Pace, R., 1995. Parametric, semiparametric, and nonparametric estimation of characteristics values within mass assessment and hedonic pricing models. J. Real. Estate Financ. Econ. 11, 195–217.

Peng, T.-C., 2019. Does the school input quality matter to nearby property prices in Taipei metropolis? An application of spatial analyses. Int. J. Hous. Mark. Anal. 12 (5), 865–883.

Preez, M., Lee, D.E., Sale, M., 2013. Nonparametric estimation of a hedonic price model: a South African case study. Stud. Econ. Econ. 37 (2), 41–62.

Qui, L., Guo, D., Zhao, X., Zhang, W., 2020. The value of school in urban China: a spatial quantile regression with housing transactions in Beijing. Appl. Econ. 52 (20), 2135–2146.

Rajapaksa, D., Gono, M., Wilson, C., Managi, S., Lee, B., Hoang, Y.-N., 2020. The demand for education: the impacts of good schools on property values in Brisbane, Australia. Land Use Policy 97, 1–7.

Rosenthal, L., 2003. The value of secondary school quality. Oxf. Bull. Econ. Stat. 65, 329–355.

Sedgley, N., Williams, N., Derrick, F., 2008. The effect of educational test scores on house prices in a model with spatial dependence. J. Hous. Econ. 17, 191–200.

Turnbull, G., Zheng, M., 2019. A meta-analysis of school quality capitalization in U.S. house prices. Real. Estate Econ. V0 (0), 1–52.

U.S. Census Bureau, 2019. Quickfacts Fremont city, California. ⟨https://www.census.gov/quickfacts/fremontcitycalifornia⟩ (Accessed 24 May 2021).

Waddell, P., Besharati-Zadeh, A., 2020. A comparison of statistical and machine learning algorithms for predicting rents in the San Francisco Bay Area. CoRR abs/2011.14924. ⟨https://arxiv.org/pdf/2011.14924.pdf⟩ (Accessed 01 October 2021).

Wen, H., Xiao, Y., Hui, E., Zhang, L., 2018. Education quality, accessibility, and housing price: Does spatial heterogeneity exist in education capitalization? Habitat Int. 78, 68–82.

Wen, H., Xiao, Y., Hui, E., 2019. Quantile effect of educational facilities on housing price: do homebuyers of higher-priced housing pay more for educational resources? Cities 90, 100–112.

Wen, H., Xiao, Y., Zhang, L., 2017. School district, education quality, and housing price: Evidence from a natural experiment in Hangzhou, China. Cities 66, 72–80.

Xiang, L., Stillwell, J., Burns, L., Heppenstall, A., Norman, P., 2018. A geodemographic classification of sub-districts to identify education inequality in central Beijing. Comput. Environ. Urban Syst. 70, 59–70.

Zahirovic-Herbert, V., Turnbull, G., 2008. School quality, house prices and liquidity. J. Real Estate Financ. Econ. 37, 113–130.

Zhang, J., Li, H., Lin, J., Zheng, W., Li, H., Chen, Z., 2020. Meta-analysis of the relationship between high quality basic education resources and housing prices. Land Use Policy 99, 1–9.

Zhang, M., Chen, J., 2018. Unequal school enrollment rights, rent yields gap, and increased inequality: the case of Shanghai. China Econ. Rev. 49, 229–240.