International Journal of
Geo-Information

MDPI

Article

The Impacts of Public Schools on Housing Prices of Residential Properties: A Case Study of Greater Sydney, Australia

Yi Lu *, Vivien Shi and Christopher James Pettit

City Futures Research Centre, School of Built Environment, University of New South Wales, Sydney 2052, Australia; ye.shi@unsw.edu.au (V.S.); c.pettit@unsw.edu.au (C.J.P.)
* Correspondence: yi.lu@unsw.edu.au

Abstract: Residential property values are influenced by a combination of physical, socio-economic and neighbourhood factors. This study investigated the influence of public schools on residential property prices. Relatively few existing models have taken the spatial heterogeneity of different submarkets into account. To fill this gap, three types of valuation models were applied to sales data from both non-strata and strata properties, and how the proximity and quality of public schools have influenced the prices of different residential property types was examined. The findings demonstrate that an increase of one unit in the normalised NAPLAN score of primary and high schools will lead to a 3.9% and 1.4%, 2.7% and 2.8% rise in housing prices for non-strata and strata properties, respectively. It is also indicated that the application of geographically weighted regression (GWR) can better capture the varying effects of schools across space. Moreover, properties located in the catchment of high-scoring schools in northern Greater Sydney are consistently the most influenced by school quality, regardless of the property type. These findings contribute to a comprehensive understanding of the relationships between public schools and the various submarkets of Greater Sydney. This is valuable for the decision-making processes of home buyers, developers and policymakers.

Keywords: automatic valuation model; housing price; public schools; geographically weighted regression; Greater Sydney

check for updates

Citation: Lu, Y.; Shi, V.; Pettit, C.J. The Impacts of Public Schools on Housing Prices of Residential Properties: A Case Study of Greater Sydney, Australia. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 298. https://doi.org/10.3390/ijgi12070298

Academic Editors: Wolfgang Kainz and Rafael Suárez-Vega

Received: 2 June 2023
Revised: 17 July 2023
Accepted: 18 July 2023
Published: 24 July 2023

1. Introduction

The total value of Australian real estate, recently estimated as A$10 trillion (Figure 1), has experienced significant growth over the past few decades [1]. Housing prices in Sydney, one of the most expensive cities in Australia, are influenced by several factors including population growth, interest rates and supply limits relative to strong demand from both local and international buyers. Various theoretical and methodological perspectives on property valuation have been explored by researchers in recent decades [2,3]. The International Association of Assessing Officers (IAAO) defines the automatic valuation model (AVM) as a mathematically based computer software program used to estimate market values. Such models are typically based on the analysis of location, market conditions and real estate characteristics. Hedonic price models (HPM), initially proposed by Rosen [4] based on the fundamentals of Lancaster's consumer theory [5], are commonly used in property valuation due to their reliability in estimating value, when assuming market equilibrium and perfect competition [6,7].

Appropriate explanatory variables and calibration methodologies are crucial to the success of hedonic price models. Researchers have investigated various hedonic variables at different spatial scales. From a broad perspective, economic and social conditions are the primary driving factors. Among many spatial variables, the proximity of a property to schools, and their quality, are important determinants, especially when the schools have outstanding reputations [8,9]. Black [10] investigated the relationship between public school quality and house prices using hedonic linear regression. The study found that parents in Massachusetts, United States were willing to pay approximately 2.1% more for houses near schools with test scores 5% higher than the mean. Kane, et al. [11] found that housing prices are influenced by both school quality and neighbourhood characteristics within school zones. Sah, et al. [12] explored the effects of proximity to schools as a control variable in the relationship between school quality and housing prices. However, with a deeper understanding of AVMs, researchers have observed that the relationship between school quality and housing prices cannot be fully captured without taking spatial effect into consideration [13].

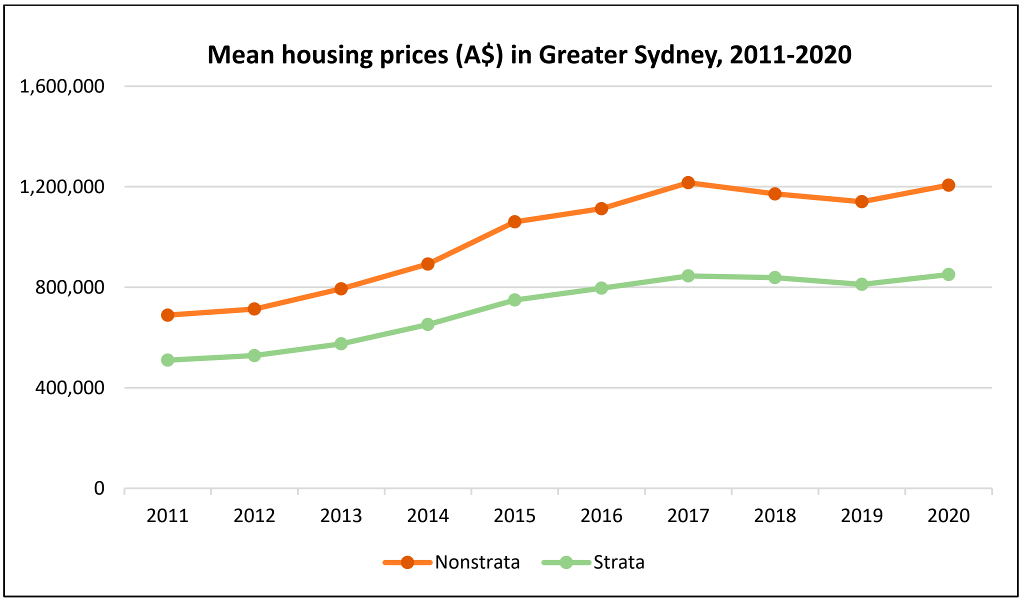

Figure 1. Mean housing prices of residential properties in Greater Sydney, 2011–2020 (Data source: Australian Property Monitors).

Fack and Grenet [14] proposed a matching framework to address fixed effects on property prices. This required that the comparison group for a transaction be in the immediate vicinity and transacted during the same school year. The outcomes of their modelling experiments demonstrated that a one-standard-deviation increase in public school performance led to a 1.4% to 2.6% rise in house prices. To control the influence of unobservable factors in Shanghai, China, Feng and Lu [15] used fixed-effects (FE) and random effects (RE) models. It was revealed that the presence of an additional Experimental Model Senior High School (the best quality school) per square kilometre led to a significant 17.1% increase in housing prices. Over the past decade, the difference-in-difference (DID) approach has been commonly used for capturing the impacts of school quality, including in Victoria, Australia [16]; Hangzhou, China [13]; and Seoul, South Korea [17]. This approach can capture variations in market response by comparing treatment and controlled areas, mitigating exogenous influences and thus yielding reliable models [18]. All of these studies found that variations in school quality influenced housing prices.

In addition to the effects of exogenous factors, it has been shown that the distribution and characteristics of house prices, just like other geographical phenomena, are always spatially diverse across submarkets [19,20]. Geographically weighted regression (GWR), which was first introduced by Brunsdon, et al. [21], has been used as a non-stationary technique. Wen, et al. [22] utilised GWR to analyse the relationship between educational facilities, their proximity and quality and housing prices. Using data from 380 counties in Poland in 2018, it is shown by the modelling results of Cellmer, et al. [23] that the impact of the analysed price determinants is spatially differentiated. Furthermore, Wang, et al. [24] proposed a regionally geographically weighted regression (RGWR) method that incorporates zoning discrimination and optimised spatial weights to improve the accuracy of geographically weighted regression (GWR) estimation and conducted an analysis of residential sale prices in Wuhan City. These studies have demonstrated that GWR can effectively model both global and local spatial relationships [25]. Nevertheless, there is still limited research examining the effects of public-school proximity and quality on the prices of various residential properties.

This paper addresses the existing research gap by investigating the spatial heterogeneity effect of public schools on housing prices. To achieve this, three alternative valuation models (AVMs) are employed: a conventional hedonic price model, a spatial lag model and a GWR-based hedonic price model. The research focuses on two distinct types of residential properties, namely strata and non-strata within Greater Sydney. Specifically, strata properties are flats, units, apartments and condominiums with strata titles, whereas non-strata properties refer to houses, semi-detached or terraces with Torrens title (also known as 'Freehold') in Australia [26]. This paper is organised as follows: Section 2 outlines the general framework and key regression methods. Section 3 presents a comprehensive case study that investigates the impact of public schools on property prices in Greater Sydney, Australia. Section 4 analyses and compares the modelling outputs and discusses the accuracy and effectiveness of regression options. Finally, Section 5 provides conclusions and limitations of the proposed study, as well as ideas for further research.

2. Materials and Methods

2.1. General Workflow

The research adopted the following workflow (Figure 2). Initial datasets were extracted from specific databases by selecting the study areas and temporal range. These datasets included proprietary sales records (Australian Property Monitors—APM) and other variables (from public and open databases) relating to structure, locational and neighbourhood. Instances with missing values or outliers, which could potentially distort the modelling process, were removed. The cleaned datasets were then used for model fitting using the different regression methods.

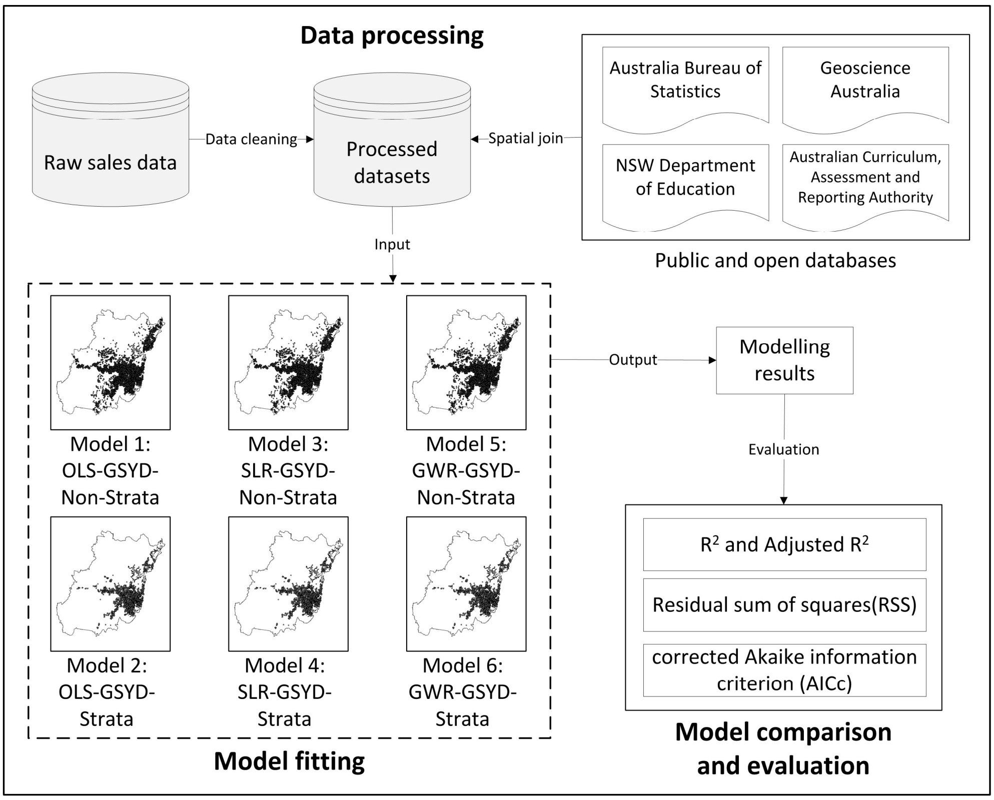

Figure 2. General framework of proposed models (OLS–Ordinary least square regression, SLR–Spatial lag regression, GWR–Geographically weighted regression).

2.2. Regression Methods

The hedonic price model (HPM) is a classic method in the field of housing market research. Ordinary least square regression (OLS) has been widely applied to estimate housing prices using specific housing characteristics as independent variables [27]. There are three groups of variables in a typical OLS model: structural (S), location (L) and neighbourhood (N) [28]. Therefore, the general form of the OLS model [29] is:

$$Log(Price) = \alpha + \sum_i \beta_i S_i + \sum_j \beta_j L_j + \sum_k \beta_k N_k + \varepsilon \tag{1}$$

where the housing price is expressed in logged form, α is a constant term, β_i refers to the coefficient for the ith variable X_i and ε is the residual error term. Additionally, the number of structural, location and neighbourhood variables are i, j and k, respectively.

The performance of OLS-based AVMs has been shown to be limited by failure to reflect the spatial dependence of housing prices on the value of neighbouring properties [30]. To accommodate this factor, spatial regression models, specifically spatial lag regression (SLR) and geographically weighted regression (GWR) models have been proposed. Amending Equation (1), the spatial lag model is formulated as:

$$Log(Price) = \alpha + \rho W y + \sum_i \beta_i S_i + \sum_j \beta_j L_j + \sum_k \beta_k N_k + \varepsilon \tag{2}$$

where ρ is an autoregressive parameter that measures the degree of spatial correlation, W is a spatial weights matrix representing the relationship between spatial units and y is a vector of observations on the random variable [31,32]. SLR improves model accuracy relative to OLS and reduces spatial bias [33].

Nevertheless, the accuracy of SLR is still limited as its regression coefficients cannot fully reflect uneven distributions of geographical features and the spatial characteristics of study area sub-regions. GWR models address this by taking the explicit locations of samples into consideration [21,34]:

$$Log(Price) = \beta_0(u_i, v_i) + \sum_i \beta_k(u_i, v_i) S_i + \sum_j \beta_j(u_i, v_i) L_j + \sum_k \beta_k(u_i, v_i) N_k + \varepsilon \tag{3}$$

where u_i and v_i are the spatial coordinates of sample i, and the remaining parameters are the same as in Equation (1).

2.3. Study Region

Greater Sydney (GSYD) is the largest metropolitan area in Australia, covering a total area of 12,368 km^2. According to the latest population census [35], Greater Sydney comprises 2,076,284 private dwellings. The influx of overseas immigrants in recent decades [36] identified as a significant factor influencing the real estate market, driving up housing prices [37]. Median housing prices have risen continuously in parallel with population growth. However, recent studies have revealed the complex and diverse nature of housing prices in GSYD [38]. As Figure 3 illustrates, the GSYD real estate market exhibits distinct regional variations, with inner Sydney emerging as the most active area for housing transactions. Given its expansive area, large population, spatial heterogeneity and consistently high demand for residential properties, GSYD is an ideal test area for the analysis.

The Australian Bureau of Statistics (ABS) aggregates census and other data at several scales. Based on the labour force, Statistical Area Level 4 (SA4) units have populations over 100,000 and may include as many as 500,000 people in metropolitan areas. There are 15 SA4s in Greater Sydney, with different housing transaction densities. To avoid the unreliable results arising from insufficient samples, SA4s with relatively low housing transaction densities (less than 5 transactions/km^2) were excluded from further analysis (Table 1).

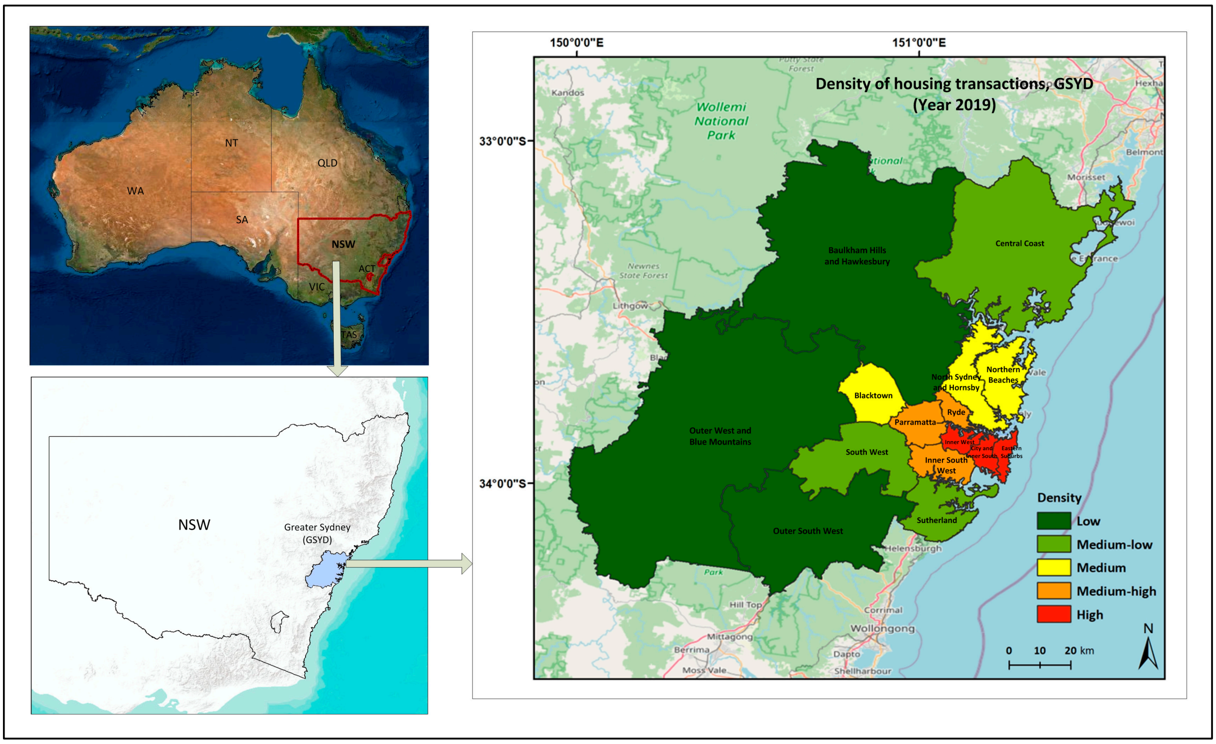

Figure 3. Location of study region and density of housing transaction (Data source: Australian Property Monitors).

Table 1. Housing transaction densities of all SA4s, Greater Sydney region.

SA4 Name	Transaction Count (Year 2019)	Area (km^2)	Transaction Density (Transactions/km^2)	Mean Housing Price (Year 2019, A$)		Note
				Non-Strata Subset	Strata Subset	
Eastern Suburbs	4144	57.73	71.78	2,885,858.75	1,211,549.64	
Inner West	4396	64.55	68.10	1,806,521.73	809,119.17	
City and Inner South	4437	66.10	67.13	1,544,313.77	975,172.73	
Inner South West	7389	163.93	45.07	1,053,518.12	582,984.94	
Ryde	3021	69.34	43.57	1,634,280.55	700,624.71	
Parramatta	6020	162.84	36.97	928,815.93	549,606.95	
North Sydney and Hornsby	7345	275.1	26.70	2,212,010.76	976,915.99	
Blacktown	5506	240.88	22.86	733,549.33	418,392.71	
Northern Beaches	4462	254.21	17.55	1,957,974.37	1,002,368.82	
Sutherland	3884	295.85	13.13	1,167,946.45	693,319.62	
South West	4579	540.28	8.48	773,194.36	412,981.55	
Central Coast	7238	1681.01	4.31	658,773.75	480,703.61	
Outer South West	4462	1277.24	3.49	655,916.57	429,996.68	
Outer West and Blue Mountains	5759	3968.13	1.45	666,902.18	389,165.66	Excluded
Baulkham Hills and Hawkesbury	3306	3251.5	1.02	1,244,046.74	734,168.11	

2.4. Data Processing and Variable Selection

Values of residential properties are mainly influenced by three groups of variables, which represent their structural (S), locational (L) and neighbourhood and environmental characteristics (N) [39,40]. Based on data availability and preliminary experiments (outlier removal and correlation testing), candidate variables were selected for the case study. As shown in Table 2, the housing prices and corresponding structural characteristics of residential properties are derived from the Australian Property Monitors (APM) in the proposed case study. Additionally, the locational (L) and neighbourhood (N) characteristics are obtained from various public and open datasets. Specifically, the logged distance to the nearest points of interest (POIs) or features of interest (FOIs) is obtained from Geoscience Australia. Furthermore, the accessibility and NAPLAN results of public schools are separately downloaded from the NSW Department of Education and the Australian Curriculum, Assessment and Reporting Authority (ACARA). Finally, the remaining neighbourhood characteristics are calculated using census data from the Australian Bureau of Statistics as the reference.

Table 2. The definition and source of selected variables.

Variable Type	Variable Name	Definition	Data Source
Dependent	Log_Price	The natural logarithm of housing price	Australian Property Monitors (APM)
Independent-Structural (S)	Bedroom	Number of bedrooms	
	Bathroom	Number of bathrooms	
	Parking	Number of carparks	
	Landsize (For non-strata subset only)	Land size	
	HasStudy (For strata subset only)	Has study room	
Independent-Locational (L)	L_CityCen	Log of distance to the nearest city centre	Geoscience Australia
	L_CoastLine	Log of distance to nearest coastline	
	L_RailSta	Log of distance to the nearest railway station	
	Near_Mainroad	Within 100 m of main roads (Yes = 1, no = 0)	
	L_Pri_Sch	Log of distance to the public primary school of the school catchment	NSW Department of Education
	L_High_Sch	Log of distance to the public high school of the school catchment	
Independent-Neighbourhood (N)	Professional_per	Percentage of professional workers	Australian Bureau of Statistics (ABS)
	Overseas_per	Percentage of residents born overseas	
	FamIncome_w	The median family income per week	
	Age65Plus_per	Percentage of residents over 65 years old	
	Prim_Ndom	Normalised National Assessment Program–Literacy and Numeracy (NAPLAN) results of year 2018 for primary school catchments	Australian Curriculum, Assessment and Reporting Authority (ACARA)
	High_Ndom	Normalised National Assessment Program–Literacy and Numeracy (NAPLAN) results of year 2018 for public high school catchments	

Non-strata property: houses, semi-detached or terraces with Torrens title (also known as 'Freehold'). Strata property: units and apartments with strata titles.

Figure 4 depicts the seven key steps involved in the data processing procedure. Initially, sales data for the second and third quarters of the year 2019 were isolated in order to reduce the impact of seasonality on data sampling [41]. The sales data comprised price, location and structural details (the variables Bedroom, Bathroom, Parking, Landsize and HasStudy in Table 3). Following this, incomplete records and price outliers, specifically the lowest 10% and highest 10%, were removed. The records were then sorted by SA4 using the ABS boundaries within GSYD. Subsequently, the primary and high school catchments within each SA4 were identified using school catchment boundaries for all New South Wales (NSW). All results from the Normalised National Assessment Program–Literacy and Numeracy (NAPLAN) were reclassified. Specifically, NAPLAN is administered to Australian students in years 3, 5, 7 and 9. Scores for Reading, Writing, Spelling, Grammar and Numeracy were divided into five groups (Well above, Above, Close to, Below, Well below) and subsequently numbered as 5 (Well above) to 1 (Well below). These numbers were then associated with the school catchments using the school name giving a normalised NAPLAN score (the mean value of Reading, Writing, Spelling, Grammar and Numeracy numbers) for each primary (Years 3 and 5) and high school (Years 7 and 9) catchment within GSYD (Figure 5). Any catchment with a null result was excluded from further analysis. The values of the locational (L) and neighbourhood (N) variables (Table 3) were estimated as the Euclidean distance between each point of interest (POI) and the sale location. Finally, a join operation was performed to connect sales records and school catchments based on their spatial relationship. This enabled us to link the NAPLAN results to all sales records within the corresponding boundaries.

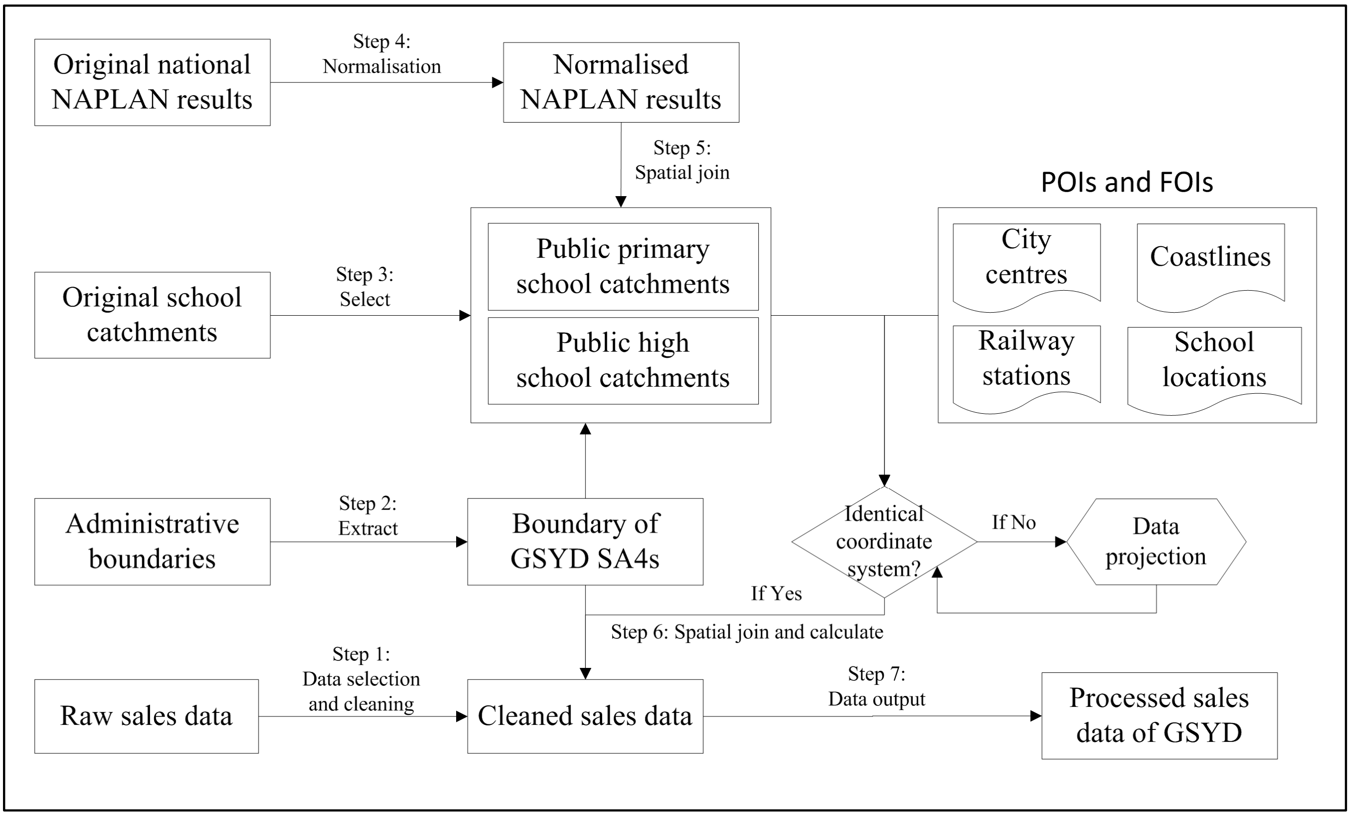

Figure 4. Workflow for data processing.

Table 3. Descriptive statistics of non-strata and strata subsets.

Variable	Non-Strata Subset Count: 13,534				Strata Subset Count: 8896			
	Min	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**
Bedroom	1.00	7.00	3.60	0.87	1.00	5.00	1.92	0.59
Bathroom	1.00	4.00	1.87	0.71	1.00	2.00	1.40	0.49
Parking	0.00	11.00	1.92	1.00	0.00	4.00	1.08	0.49
Landsize	42.64	23,868.28	801.34	1178.12	-	-	-	-
HasStudy	-	-	-	-	0.00	1.00	0.17	0.38
L_CityCen	6.10	10.34	9.12	0.59	4.29	10.32	8.83	0.81
L_CoastLine	4.54	10.77	9.47	0.98	3.47	10.58	8.83	1.22
L_Pri_Sch	2.38	8.41	6.25	0.59	3.16	7.61	6.08	0.62
L_High_Sch	2.84	9.18	6.87	0.65	1.71	9.10	6.61	0.73
L_RailSta	3.90	9.69	7.48	0.88	3.56	9.69	6.79	1.18
Near_Mainroad	0.00	1.00	0.38	0.48	0.00	1.00	0.63	0.48
Professional_per	0.00	50.12	19.82	8.03	0.00	61.29	26.12	9.48
Overseas_per	5.85	89.46	36.93	13.20	0.00	94.44	53.70	17.56
FamIncome_w	754.00	5250.00	2228.43	626.95	0.00	5250.00	2191.58	634.55
Age65Plus_per	0.00	93.53	12.63	7.24	0.00	87.12	8.86	7.48
Prim_Ndom	1.00	5.00	3.39	1.20	1.00	5.00	3.85	1.14
High_Ndom	1.00	5.00	3.02	1.21	1.00	5.00	3.43	1.06

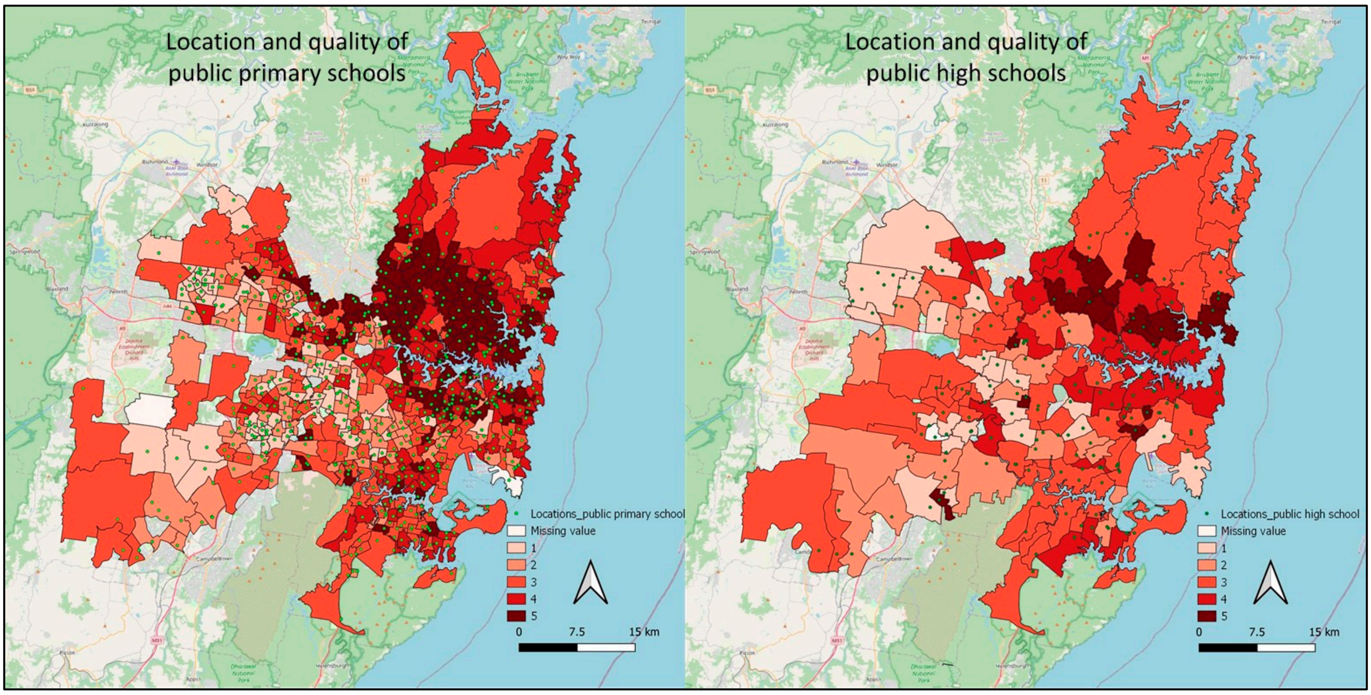

Figure 5. Locations and quality of public schools.

To understand the influence of variables in different submarkets, the entire dataset was divided into two subsets (non-strata and strata). In the second and third quarters of 2019, there were 13,534 non-strata sales and 8896 strata sales within the included SA4s. The descriptive statistics of the two subsets are in Table 3.

3. Results

Table 4 shows the results of the OLS and SLR models, with logarithm of the sales price as the dependent variable, in each of the residential property groups. The adjusted R2 values of models 1 to 4 are 0.701, 0.608, 0.701 and 0.609, respectively. Thus, the non-strata models explain more than 70% of the variance in the natural logarithm of housing price, while the strata models explain more than 60% of the variance. Each of the structural (S) and neighbourhood (N) variables has a positive influence on housing prices. Among the locational (L) variables, the logarithms of distance to the nearest city centres (L_CityCen), to the coastline (L_CoastLine), to primary school (L_Pri_Sch) and to railway stations (L_RailSta) indicate that proximity has a positive influence, regardless of property type. However, residential properties located near high schools near main roads (Near_Mainroad) have reduced prices.

Subsequently, GWR-based models were applied to examine the impact of public schools while taking spatial heterogeneity into account. The GWR-based models were developed using the mgwr 2.1.1 package [42] in Python 3.7. The pre-processed sales data were imported and the optimal bandwidths and number of neighbours for each local regression, were calculated. Using a bi-square kernel function, the optimal bandwidths are determined as 374.0 and 254.0 for the non-strata and strata subsets. Afterwards, model fitting is implemented with these calibrated bandwidths, and the modelling results are recorded in Table 5. The adjusted R2 values were 0.855 for non-strata and 0.817 for strata prices, indicating that the GWR models better explained price variance than either the OLS or SLR models. Moreover, the relatively lower values of RSS, AICc and Moran's I of residuals also demonstrated that there was less unexplained variation in the results of models 5 and 6. The coefficients of the school-related variables are presented in Figures 6 and 7 (non-strata and strata subsets, separately).

Table 4. The results of ordinary and spatial linear regression models.

	Variable	**Model 1: OLS-Non-Strata**	**Model 2: OLS-Strata**	**Model 3: SLR-Non-Strata**	**Model 4: SLR-Strata**
	Constant	14.893 ***	13.531 ***	14.909 ***	13.513 ***
Independent-Structural (S)	Bedroom	0.075 ***	0.159 ***	0.075 ***	0.160 ***
	Bathroom	0.048 ***	0.122 ***	0.048 ***	0.121 ***
	Parking	0.034 ***	0.043 ***	0.034 ***	0.044 ***
	Landsize	0.000 ***	–	0.000 ***	–
	HasStudy	–	0.054 ***	–	0.054 ***
Independent-Locational (L)	L_CityCen	−0.059 ***	−0.036 ***	−0.058 ***	−0.036 ***
	L_CoastLine	−0.162 ***	−0.100 ***	−0.161 ***	−0.099 ***
	L_Pri_Sch	−0.026 ***	−0.010 ***	−0.027 ***	−0.010 ***
	L_High_Sch	0.018 ***	0.014 ***	0.017 ***	0.014 ***
	L_RailSta	−0.026 ***	−0.021 ***	−0.027 ***	−0.020 ***
	Near_Mainroad	−0.020 ***	−0.012 **	−0.020 ***	−0.010 **
Independent-Neighbourhood (N)	Professional_per	0.010 ***	0.007 ***	0.011 ***	0.006 ***
	Overseas_per	0.004 ***	0.003 ***	0.004 ***	0.002 ***
	FamIncome_w	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	Age65Plus_per	0.008 ***	0.006 ***	0.008 ***	0.006 ***
	Prim_Ndom	0.039 ***	0.014 ***	0.039 ***	0.014 ***
	High_Ndom	0.027 ***	0.029 ***	0.028 ***	0.029 ***
	W_Log_Price			−0.003 ***	0.003 ***
Modelling result	Observations	13,534	8896	13,534	8896
	Adjusted R^2/Spatial Pseudo R^2	0.701	0.608	0.701	0.609
	Residual sum of squares (RSS)	508.841	227.603	508.666	226.959
	AICc	−5958.807	−7328.704		
	Moran's I of residuals	0.755 (p = 0)	0.707 (p = 0)	0.755 (p = 0)	0.702 (p = 0)

Significance levels: ** $p < 0.01$, *** $p < 0.001$.

Table 5. The coefficients of GWR-based modelling results.

Variable	Model 5: GWR-GSYD-Non-Strata			Model 6: GWR-GSYD-Strata		
	Mean	**Min**	**Max**	**Mean**	**Min**	**Max**
Constant	17.313	−2034.722	1107.776	6.899	−1515.958	1493.882
Bedroom	0.078	−0.004	0.199	0.169	−0.003	0.332
Bathroom	0.055	−0.108	0.160	0.126	−0.016	0.328
Parking	0.027	−0.055	0.095	0.070	−0.019	0.177
Landsize	0.000	0.000	0.001	-	-	-
HasStudy	-	-	-	0.038	−0.078	0.155
L_CityCen	−0.146	−4.870	5.158	0.143	−3.653	11.883
L_CoastLine	−0.323	−12.972	5.722	0.293	−6.883	27.178
L_Pri_Sch	−0.003	−0.131	0.122	−0.006	−0.193	0.113
L_High_Sch	0.007	−0.121	0.138	0.012	−0.194	0.856
L_RailSta	−0.011	−0.846	1.011	0.060	−1.305	4.372
Near_Mainroad	−0.029	−0.160	0.190	−0.022	−0.135	0.071
Professional_per	0.001	−0.045	0.029	0.002	−0.021	0.019
Overseas_per	0.000	−0.014	0.013	0.000	−0.008	0.012
FamIncome_w	0.000	0.000	0.001	0.000	0.000	0.001
Age65Plus_per	0.002	0.010	0.024	0.003	−0.013	0.024
Prim_Ndom	0.002	−0.321	0.192	0.026	−90.803	42.177
High_Ndom	0.210	−364.997	693.184	0.264	−368.589	465.801
Observations		13,534			8896	
Adjusted R^2		0.855			0.817	
Residual sum of squares (RSS)		222.708			94.991	
AICc		−14,300.532			−13,058.085	
Moran's I of residuals		0.593 (p = 0)			0.347 (p = 0)	

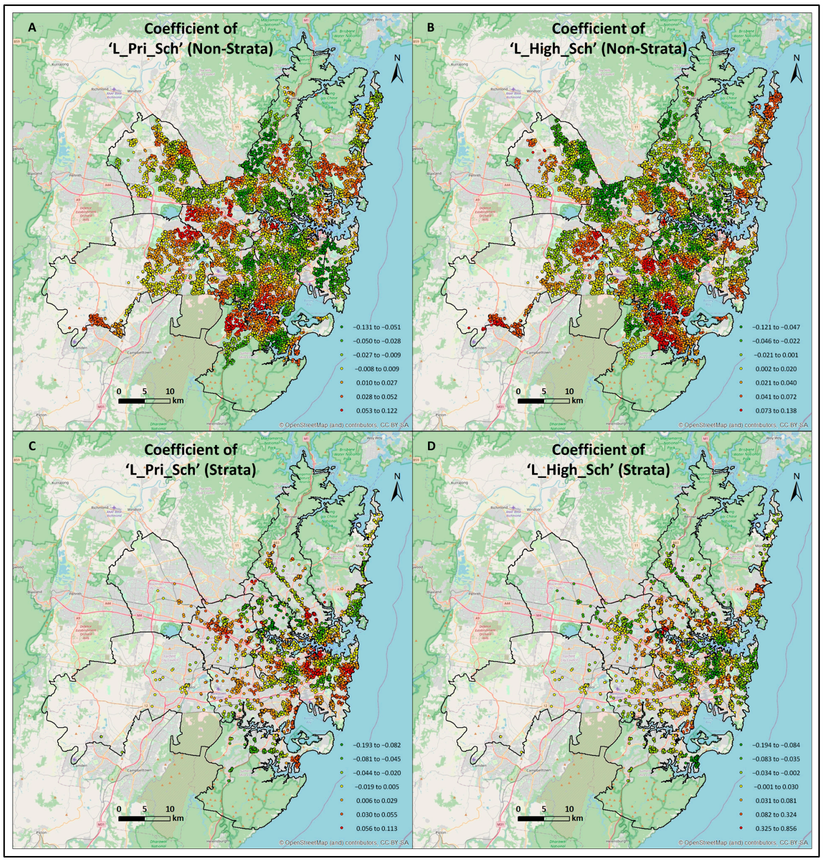

Figure 6. Coefficient distribution of school proximity variables. Variables 'L_Pri_Sch' and 'L_High_Sch' represent the log of distance to the public primary and high schools of the school catchments. The green (negative) values mean higher prices nearer to the school.

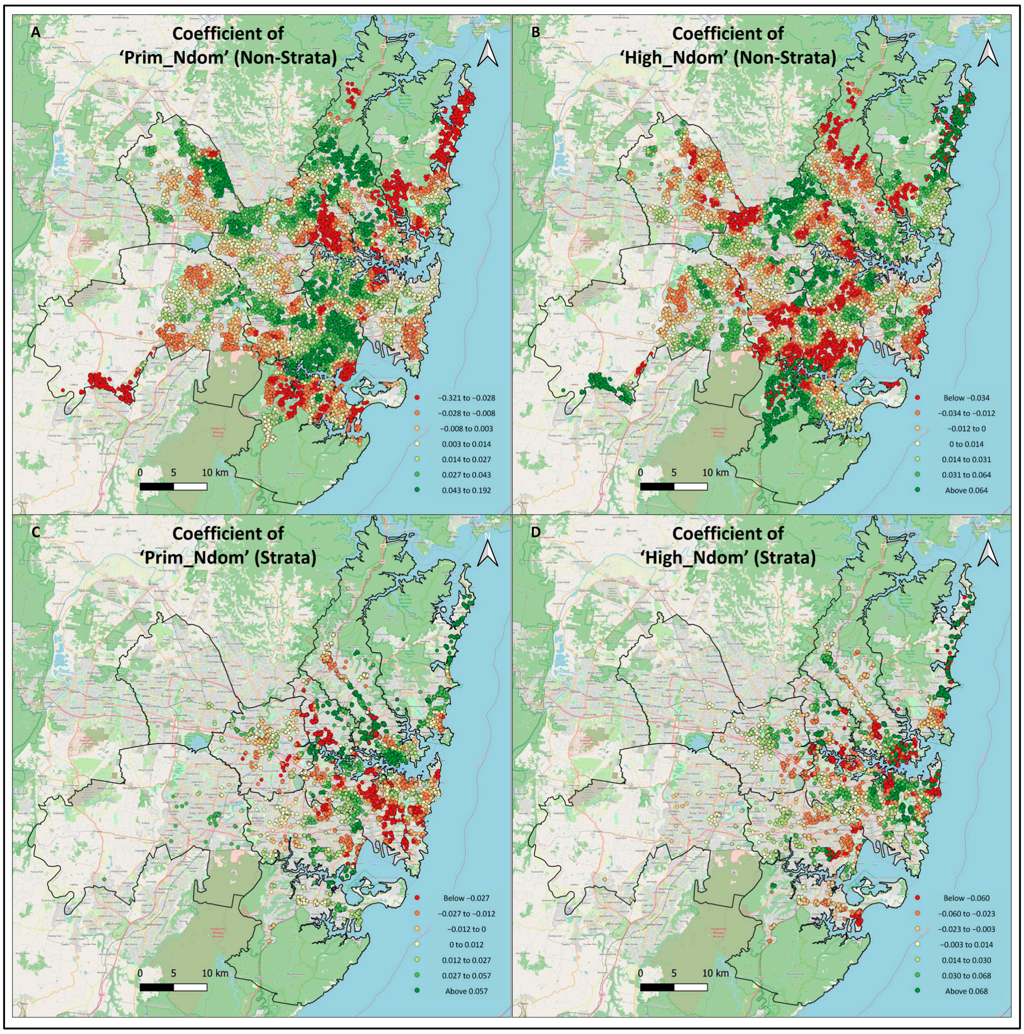

Figure 7. Coefficient distribution of school quality variables. Variables 'Prim_Ndom' and 'High_Ndom' represent the normalised NAPLAN results for public primary and high school catchments. The green (positive) values indicate higher prices in high-quality school areas.

4. Discussion

The results of OLS and SLR models demonstrate a strong association between the proximity and quality of public schools and housing prices in Greater Sydney. The modelling of the non-strata subset showed that a one-point rise in the normalised NAPLAN score for public primary schools resulted in a 3.9% increase in the natural logarithm of housing prices. Similarly, for high schools, a one-point rise in normalised NAPLAN score leads to a 2.7% (model 1) or 2.8% (model 3) increase in the natural logarithm of housing prices. However, the relationship with proximity is more nuanced, as housing prices were found to be higher near primary schools but lower near high schools. Specifically, a one-unit decrease in the logged distance to a primary school results in a 2.6% (model 1) and 2.7% (model 3) increase in the natural logarithm of housing prices. The strata subset (models 2 and 4) had similar results, with coefficients of 1.4%, 2.9% and −1.0% for the 'Prim_Ndom', 'High_Ndom' and 'L_Pri_Sch' variables, respectively. These overall findings are consistent with previous case studies, which found that proximity to and high quality of public schools have a positive impact on the values of residential properties [43,44]. Nevertheless, in the case of Greater Sydney, there is an exception with the negative influence of the proximity to high schools.

Similar to the findings of OLS and SLR models, GWR-based models have also confirmed that the quality of public schools significantly affects the housing prices of residential properties, regardless of their type. In terms of non-strata subset, properties located within

Greater Sydney can receive an increase in the natural algorithm of housing prices of up to 19.2% for each additional normalised NAPLAN score for public primary schools and 27.4% for public high schools (Figure 7A,B), excluding extreme values as shown in Table 5. The North Sydney and Hornsby, Inner South West, Blacktown and Inner West SA4 regions exhibit a relatively high mean coefficient of the 'Prim_Ndom' variable (Table 6). Firstly, the northern part of Greater Sydney is known for its superior public education resources, and residents highly value the assessment results of public schools in their school catchments. Similarly, some primary school catchments in the western part of Greater Sydney, such as Ashfield and Burwood Public Schools, are also recognised for exceptional academic performances. Therefore, higher mean coefficients of 'Prim_Ndom' is observed within the afore-mentioned parts of Greater Sydney. Additionally, the SA4s of Sutherland, Northern Beaches and Inner South West have shown higher mean coefficients of the 'High_Ndom' variable. It is also likely to be led by the availability of exceptional public high school resources in these regions. Furthermore, with the increased age of students and the well-developed public transportation system in Greater Sydney, the proximity to public high schools is less critical than the proximity to public primary schools (Figure 6A,B). This finding aligns with the conclusions from other researchers that the accessibility effect of primary schools on housing price can be more significant [45] and also explains why there are more samples (n = 7114) with negative coefficients of the variable 'L_Pri_Sch' than the number of samples (n = 5543) with negative coefficients of the variable 'L_High_Sch'.

Table 6. The mean value of quality-related coefficients in all SA4s.

SA4 Name	Non-Strata Subset (Model 5)		Strata Subset (Model 6)	
	Variable 'Prim_Ndom'	Variable 'High_Ndom'	Variable 'Prim_Ndom'	Variable 'High_Ndom'
Blacktown	0.023	−0.004	0.018	0.006
City and Inner South	0.009	0.015	−0.019	0.032
Eastern Suburbs	0.000	−0.006	−0.013	0.163
Inner South West	0.024	0.315	0.014	0.082
Inner West	0.023	0.023	0.021	−0.017
North Sydney and Hornsby	0.034	0.001	0.121	0.071
Northern Beaches	−0.089	0.510	0.035	2.654
Parramatta	0.017	0.012	0.024	−0.001
Ryde	−0.020	0.020	−0.031	0.019
South West	−0.023	0.043	0.028	0.007
Sutherland	−0.011	1.059	0.017	−0.024

Regarding the strata subset, 'Prim_Ndom' with the highest mean coefficient value is observed in North Sydney and Hornsby SA4s (Table 6). Nonetheless, there is a discernible difference between models 5 and 6 as the overall mean coefficient of the 'Prim_Ndom' variable is noticeably larger in the strata subset (0.026) than in the non-strata subset (0.002) (Table 5). This specific point of modelling result suggests that the quality of public primary schools has a greater impact on housing prices in the majority of strata properties than non-strata properties. On the other hand, the coefficient distribution for the 'Prim_Ndom' and 'High_Ndom' variables (Figure 7C,D) also indicates that the quality of public high schools becomes even more significant than that of public primary schools in the SA4s of City and Inner South and Eastern Suburbs. These modelling results have been compared with the ABS 2016 Index of Relative Socio-Economic Disadvantage (IRSD), which indicates the relative level of socio-economic disadvantage [46]. Residents living in SA4s with higher IRSD scores (i.e., low disadvantage) are more inclined to choose private or elite primary schools in order to gain access to high-quality public high schools. Therefore, the quality of a public primary school appears to produce less influence than that of a public high school. This also illustrates the distribution of variables 'L_Pri_Sch' (Figure 6C) and 'L_High_Sch' (Figure 6D) by highlighting the influence of proximity to public high schools, relative to public primary schools, on prices of strata properties in the inner-city, northern, and eastern Greater Sydney regions.

5. Conclusions

This research examines the relationship between housing prices and various variables, including proximity to and quality of public schools, through a case study in Greater Sydney. Three hedonic price models, employing OLS, SLR and GWR regression, have been tested to provide a systematic analysis of their impact. The influence of public primary school quality is higher for non-strata properties than strata properties, but for public high schools, the strata prices are more affected. Specifically, an increase of one unit in the normalised NAPLAN score of public primary schools leads to a 3.9% rise in the natural logarithm of housing prices for non-strata properties. For strata properties, there is a 1.4% increase in the natural logarithm of housing prices with the same increase. Regarding the NAPLAN-based high school quality, these rates decline to 2.7% for non-strata properties but increase to 2.8% for strata properties. However, the coefficients generated from the whole area model do not capture the impact of the unequally distributed educational resources on various submarkets across the study region.

With the aid of GWR-based hedonic price models, the localised influence of the school-related variables was explored. The GWR-based models have generated more accurate results with less spatial autocorrelation than OLS or SLR in the prediction residuals, which is in line with the prior findings [47]. Non-strata properties located in specific northern and western regions of Greater Sydney are the most influenced by the quality of both public primary and high schools. Moreover, accessibility to public primary schools appears more important than accessibility to public high schools. Regarding the strata subset, public primary school quality is still considered an important factor, especially for strata properties in northern Greater Sydney. In contrast, the positive coefficient for public high school quality is more widespread and includes the SA4s of City and Inner South and Eastern Suburbs. The influence of proximity to public primary and high schools varies substantially across the strata subset, with both high and low (negative) values of the coefficients observed in almost every SA4. It, therefore, appears that accessibility is not as significant as school quality in Greater Sydney. This accords with other studies in Melbourne [16] and Brisbane [48], the second- and third-largest Australian metropolitan areas.

It is acknowledged that the NAPLAN assessment is not a direct measure of overall school quality. Nevertheless, it is still widely accepted as an important indicator for monitoring and evaluating the performance of schools at a national level and has been utilised in other Australian-based AVMs [16,49]. Considering the distinctive characteristics of public and private schools, investigating the impact of both types on housing prices is equally important. Consequently, future research will focus on examining the influence of both public and private schools on non-strata and strata property prices in different sub-regions. Furthermore, future research will also explore the integration of AVMs with AI and machine learning methods, which are anticipated to better identify complex patterns of property valuation within sales data [50,51], and improve the accuracy of AVMs in a further step.

Author Contributions: Conceptualisation, Yi Lu and Vivien Shi; Data curation, Yi Lu and Vivien Shi; Formal analysis, Yi Lu; Methodology, Yi Lu and Vivien Shi; Resources, Christopher James Pettit; Supervision, Christopher James Pettit; Validation, Yi Lu and Vivien Shi; Visualisation, Yi Lu, Vivien Shi and Christopher James Pettit; Writing—original draft, Yi Lu; Writing—review and editing, Yi Lu, Vivien Shi and Christopher James Pettit. All authors have read and agreed to the published version of the manuscript.

Funding: The authors disclose receipt of the following financial support for the research and authorship of this article: This work has been supported by FrontierSI, a not-for-profit company that exists to deliver major benefits to governments, industry and the community in Australia and New Zealand through the application of spatial information. This research was funded through the Cooperative Research Centre Project—Value Australia (RG192482).

Data Availability Statement: The data presented in this study are subject to third-party restrictions. The Property data are available from https://www.apm.com.au/ (accessed on 18 May 2023) with the permission of APM.

Acknowledgments: The authors would like to thank Ian Bishop, an Honorary Professorial Fellow in the Department of Infrastructure Engineering at The University of Melbourne, for his proofreading and revision suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. ABS. Value of Residential Dwellings Passes $10 Trillion. Available online: https://www.abs.gov.au/media-centre/media-releases/value-residential-dwellings-passes-10-trillion (accessed on 26 February 2023).
2. Panduro, T.E.; Veie, K.L. Classification and valuation of urban green spaces—A hedonic house price valuation. *Landsc. Urban Plan.* **2013**, *120*, 119–128. [CrossRef]
3. Iban, M.C. An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat Int.* **2022**, *128*, 102660. [CrossRef]
4. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Politisk Econ.* **1974**, *82*, 34–55. [CrossRef]
5. Lancaster, K.J. A New Approach to Consumer Theory. *J. Political Econ.* **1966**, *74*, 132–157. [CrossRef]
6. Chau, K.W.; Chin, T. A critical review of literature on the hedonic price model. *Int. J. Hous. Sci. Its Appl.* **2003**, *27*, 145–165.
7. Zhu, J.; Pawson, H.; Han, H.; Li, B. How can spatial planning influence housing market dynamics in a pro-growth planning regime? A case study of Shanghai. *Land Use Policy* **2022**, *116*, 106066. [CrossRef]
8. Haurin, D.R.; Brasington, D. School Quality and Real House Prices: Inter- and Intrametropolitan Effects. *J. Hous. Econ.* **1996**, *5*, 351–368. [CrossRef]
9. Irwin, N.B.; Livy, M.R. Days and Confused: Housing Price and Liquidity Response to New Local Public Schools. *J. Real Estate Res.* **2021**, *43*, 21–46. [CrossRef]
10. Black, S.E. Do Better Schools Matter? Parental Valuation of Elementary Education. *Q. J. Econ.* **1999**, *114*, 577–599. [CrossRef]
11. Kane, T.J.; Riegg, S.K.; Staiger, D.O. School quality, neighborhoods, and housing prices. *Am. Law Econ. Rev.* **2006**, *8*, 183–212. [CrossRef]
12. Sah, V.; Conroy, S.J.; Narwold, A. Estimating School Proximity Effects on Housing Prices: The Importance of Robust Spatial Controls in Hedonic Estimations. *J. Real Estate Financ. Econ.* **2015**, *53*, 50–76. [CrossRef]
13. Wen, H.; Xiao, Y.; Zhang, L. School district, education quality, and housing price: Evidence from a natural experiment in Hangzhou, China. *Cities* **2017**, *66*, 72–80. [CrossRef]
14. Fack, G.; Grenet, J. When do better schools raise housing prices? Evidence from Paris public and private schools. *J. Public Econ.* **2010**, *94*, 59–77. [CrossRef]
15. Feng, H.; Lu, M. School quality and housing prices: Empirical evidence from a natural experiment in Shanghai, China. *J. Hous. Econ.* **2013**, *22*, 291–307. [CrossRef]
16. Haisken-DeNew, J.; Hasan, S.; Jha, N.; Sinning, M. Unawareness and selective disclosure: The effect of school quality information on property prices. *J. Econ. Behav. Organ.* **2018**, *145*, 449–464. [CrossRef]
17. Park, H.; Tidwell, A.; Yun, S.; Jin, C. Does school choice program affect local housing prices?: Inter-vs. intra-district choice program. *Cities* **2021**, *115*, 103237. [CrossRef]
18. Koo, K.M.; Liang, J. The Effect of Bilingual Education on Housing Price-a Case Study of Bilingual School Conversion. *J. Real Estate Financ. Econ.* **2020**, *62*, 629–664. [CrossRef]
19. Helbich, M.; Brunauer, W.; Vaz, E.; Nijkamp, P. Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria. *Urban Stud.* **2013**, *51*, 390–411. [CrossRef]
20. Wu, Y.; Wei, Y.D.; Li, H. Analyzing Spatial Heterogeneity of Housing Prices Using Large Datasets. *Appl. Spat. Anal. Policy* **2019**, *13*, 223–256. [CrossRef]
21. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [CrossRef]
22. Wen, H.; Xiao, Y.; Hui, E.C.; Zhang, L. Education quality, accessibility, and housing price: Does spatial heterogeneity exist in education capitalization? *Habitat Int.* **2018**, *78*, 68–82. [CrossRef]
23. Cellmer, R.; Cichulska, A.; Bełej, M. Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 380. [CrossRef]

24. Wang, Z.; Zhao, Y.; Zhang, F. Simulating the Spatial Heterogeneity of Housing Prices in Wuhan, China, by Regionally Geographically Weighted Regression. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 129. [CrossRef]
25. Yao, J.; Fotheringham, A.S. Local Spatiotemporal Modeling of House Prices: A Mixed Model Approach. *Prof. Geogr.* **2015**, *68*, 189–201. [CrossRef]
26. Gitelman, E.; Otto, G. Supply Elasticity Estimates for the Sydney Housing Market. *Aust. Econ. Rev.* **2012**, *45*, 176–190. [CrossRef]
27. Owusu-Ansah, A. A review of hedonic pricing models in housing research. *J. Int. Real Estate Constr. Stud.* **2011**, *1*, 19.
28. Mok, H.M.K.; Chan, P.P.K.; Cho, Y.-S. A hedonic price model for private properties in Hong Kong. *J. Real Estate Financ. Econ.* **1995**, *10*, 37–48. [CrossRef]
29. Zietz, J.; Zietz, E.N.; Sirmans, G.S. Determinants of House Prices: A Quantile Regression Approach. *J. Real Estate Financ. Econ.* **2007**, *37*, 317–333. [CrossRef]
30. Pace, R.K.; LeSage, J.P. Omitted variable biases of OLS and spatial lag models. In *Progress in Spatial Analysis*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 17–28.
31. Li, H.; Wei, Y.D.; Yu, Z.; Tian, G. Amenity, accessibility and housing values in metropolitan USA: A study of Salt Lake County, Utah. *Cities* **2016**, *59*, 113–125. [CrossRef]
32. Anselin, L. Spatial econometrics. In *A Companion to Theoretical Econometrics*; Blackwell: Oxford, UK, 2001; p. 310330.
33. Farber, S.; Yeates, M. A comparison of localized regression models in a hedonic house price context. *Can. J. Reg. Sci.* **2006**, *29*, 405–420.
34. Fotheringham, A.S.; Brunsdon, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
35. ABS. *Greater Sydney—2021 Census All Persons QuickStats*; Australian Bureau of Statistics: Canberra, Australia, 2021.
36. Bangura, M.; Lee, C.L. Spatial connectivity and house price diffusion: The case of Greater Sydney and the regional cities and centres of new south wales (NSW) in Australia. *Habitat Int.* **2023**, *132*, 102740. [CrossRef]
37. Pavlov, A.; Somerville, T. Immigration, Capital Flows and Housing Prices. *Real Estate Econ.* **2020**, *48*, 915–949. [CrossRef]
38. Thackway, W.T.; Ng, M.K.M.; Lee, C.-L.; Shi, V.; Pettit, C.J. Spatial variability of the 'Airbnb effect': A spatially explicit analysis of Airbnb's impact on housing prices in Sydney. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 65. [CrossRef]
39. Wen, H.; Zhang, Y.; Zhang, L. Assessing amenity effects of urban landscapes on housing price in Hangzhou, China. *Urban For. Urban Green.* **2015**, *14*, 1017–1026. [CrossRef]
40. Calka, B. Estimating Residential Property Values on the Basis of Clustering and Geostatistics. *Geosciences* **2019**, *9*, 143. [CrossRef]
41. Kaplanski, G.; Levy, H. Real estate prices: An international study of seasonality's sentiment effect. *J. Empir. Financ.* **2012**, *19*, 123–146. [CrossRef]
42. Oshan, T.M.; Li, Z.; Kang, W.; Wolf, L.J.; Fotheringham, A.S. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 269. [CrossRef]
43. Pettit, C.; Shi, Y.; Han, H.; Rittenbruch, M.; Foth, M.; Lieske, S.; van den Nouwelant, R.; Mitchell, P.; Leao, S.; Christensen, B. A new toolkit for land value analysis and scenario planning. *Environ. Plan. B Urban Anal. City Sci.* **2020**, *47*, 1490–1507. [CrossRef]
44. Yang, L.; Zhang, S.; Guan, M.; Cao, J.; Zhang, B. An Assessment of the Accessibility of Multiple Public Service Facilities and Its Correlation with Housing Prices Using an Improved 2SFCA Method—A Case Study of Jinan City, China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 414. [CrossRef]
45. Wen, H.; Zhang, Y.; Zhang, L. Do educational facilities affect housing price? An empirical study in Hangzhou, China. *Habitat Int.* **2014**, *42*, 155–163. [CrossRef]
46. ABS. Socio-Economic Indexes for Areas (SEIFA) 2016. Available online: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 (accessed on 26 February 2023).
47. Chen, S.; Zhuang, D.; Zhang, H. GIS-Based Spatial Autocorrelation Analysis of Housing Prices Oriented towards a View of Spatiotemporal Homogeneity and Nonstationarity: A Case Study of Guangzhou, China. *Complexity* **2020**, *2020*, 1079024. [CrossRef]
48. Rajapaksa, D.; Gono, M.; Wilson, C.; Managi, S.; Lee, B.; Hoang, V.-N. The demand for education: The impacts of good schools on property values in Brisbane, Australia. *Land Use Policy* **2020**, *97*, 104748. [CrossRef]
49. Lee, H.; Han, H.; Pettit, C.; Gao, Q.; Shi, V. Machine learning approach to residential valuation: A convolutional neural network model for geographic variation. *Ann. Reg. Sci.* **2023**, 1–21. [CrossRef]
50. Morano, P.; Tajani, F.; Torre, C.M. Artificial intelligence in property valuations. An application of artificial neural networks to housing appraisal. In *Advances in Environmental Science and Energy Planning*; WSEAS Press: Athens, Greece, 2015; pp. 23–29.
51. Xu, X.; Zhang, Y. Retail Property Price Index Forecasting through Neural Networks. *J. Real Estate Portf. Manag.* **2022**, *29*, 1–28. [CrossRef]