Spatial effects of public service accessibility on house prices: A case study of Shenzhen, GBA, China

Hongjing Xiao

August 2021

This dissertation is submitted in part requirement for the MSc in Smart Cities and Urban Analytics at the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL.

Submitted: 23 August 2021
Supervisor: Dr Duncan Smith
Module ID: CASA0010
Student ID: 20156725
Word count:9276

Abstract

Public service accessibility is an important reflection of the degree of economic prosperity, infrastructural development and service frequency of a city. As an important indicator for citizens' daily lives, more and more studies turn their focus to the relationship between accessibility and residential housing. Many utilizes the traditional hedonic pricing method and incorporate accessibility to public services as neighborhood characteristics, in order to understand how accessibility may affect households' willingness to purchase a property. However, modelling the relationships between accessibility and house prices is not always easy in an urban area. The intrinsic spatial heterogeneity and spatial autocorrelation, namely spatial effects seriously affect the results of traditional models (e.g. OLS model). After a series of justified usage, it is recognized by many scholars that Geographically Weighted Regression (GWR) is superior in dealing with spatial effects within the ground of hedonic pricing model, and has become a popular tool in modelling the relationship between service accessibility and house price. This study therefore adopts the long-justified hedonic pricing model together with GWR technique to model the spatial effects of public service accessibility on house price, taking Shenzhen as the sample city, and uses more recent data (collected for the year 2020). Within Shenzhen's highly segregated housing market, it is found that public service accessibility has positive influence on house prices in neighborhoods, especially those lacking accessibility to public services. Meanwhile, it is observed that accessibility to certain public service (e.g. schools) maintains strong influence on house prices across the whole city, while housing purchasers in certain areas do have specific preference of service. The usage of GWR and hedonic pricing structure in Shenzhen again prove the superiority of the two methods, and provide new insights on Shenzhen's housing market.

Declaration

I, Hongjing Xiao, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 9276 words in length.

Signed:

Date: 23/08/2021

Table of contents

List of figures

List of Tables

List of Acronyms and Abbreviations

SZ	Shenzhen
GBA	Greater Bay Area
SEZ	Special Economic Zone
POI	Points of Interest
SAC	Spatial Autocorrelation
SH	Spatial Heterogeneity
WTP	Willingness to Pay
OLS	Ordinary Least Squares
GWR	Geographically Weighted Regression
CV	Cross Validation
AIC	Akaike Information Criterion
MI	Moran’s I Index
VIF	Variance Inflation Factors
EDU	Education facilities
FOOD	Food facilities
SHOP	Shop facilities
LEIS	Leisure facilities
BANK	Financial facilities
MALL	Shopping malsl
BLD	Skyscrapers/High-rise buildings
GOV	Governmental facilities
TOUR	Tourist spots
METRO	Metro stations
BUS	Bus stations
CBD	Central Business District
GREEN	Green space

Acknowledgements

First of all, I am hugely thankful for my supervisor Duncan Smith for his guidance and patience throughout the dissertation writing process. His supervision help shape this research and overcome its shortcomings.

I am very grateful to Dr Qili Gao at Shenzhen University and Bowen Zhang at Kings College London, for their wonderful support in finding data sources for this research, and in the process of which I learned a lot concerning data processing.

I am very thankful for the Centre for Advanced Spatial Analysis (CASA) at University College London for giving me this opportunity to learn and plan a project. The skills and literacy I learned throughout the study period are prerequisites for shaping this research.

Finally, I would like to thank my parents for their thoughtful support and encouragement during the dissertation writing process. Thank you very much for the everyday companion.

Chapter 1 Introduction

1.1 Context and Motivation

Spatial effects of accessibility have long been studied ever since the neoclassical microeconomic theory, in which the ‘access-space’ trade-off is emphasized as one fundamental urban law (Alonso, 1964). In this equilibrium, higher travel cost balance lower land cost in less accessible part of a city, which gives a sense that such insight can be fitted into housing research, in particular studying the spatial effects of accessibility on house prices. From a today’ view, Alonso’s (1964) monocentric model dose not adequately capture the spatial effects of accessibility, as in the monocentric model, destinations, namely workplaces are located in the city center, while in reality that is not solely the case (McDonald, 1987). The multicentric nature of housing-market areas contains multiple dimensions that are required to be taken into account. That nature motivates upcoming researches to find alternative methods to capture how spatial effects of accessibility work on balancing house prices, which certainly goes beyond measuring physical distance and travel cost with distinctly defined urban centers.

As more and more studies on hedonic pricing theory reveal that housing and housing market is not a homogenous goods and service, but one comprising various characteristics, the concept of accessibility also find its place as utility attributes (Rosen, 1974). Beckmann and Martin (1973) made a central assumption that housing and accessibility are jointly purchased in a residential choice decision. Michelson (1977) in his discussion of individual evaluation on residential purchase, found that quality of housing mattered more for suburban dwellers while distance took priority for those in the city, indicating valuation of accessibility varied over space. In a more recent study by Evans (1995), it was argued that hedonic models developed the trade-off models with extended determinants of price from simply location-based to a range of location and dwelling characteristics. That is how accessibility-space literature has been linked to hedonic price modelling. Moreover, two ways to define access are proposed by Gibbons and Machin (2005): distance to the target and service frequency at the nearest targets. Such definition not only provides flexibility in choosing techniques to study accessibility within hedonic pricing, but also shine the path to explore further external effects such as being accessible to urban public services, and examine how they may impact house prices.

Urban public service facilities are basic factors affecting the living quality of residents and indices of attractiveness for different locations. A wide range of studies have been carried out towards the external effects brought by public service resources to housing prices. Focuses of these studies has been education, green space, subways and other public service facilities that are related to citizens' daily lives. Many found that accessibility towards certain public service (e.g. schools and education) positively impact housing prices (Lan et al., 2018). Whereas in these studies, two main methods are used to quantify the accessibility of a service: one is using a combination of distance, commuting time or expense of a mean of transport to represent accessibility (Mok and Ling-Hin, 2010), another is measuring accessibility by the availability or proximity of relative resources within a given distance (Wen et al., 2018). There are also studies asserted that certain service does not significantly affect housing price, indicating that every time the scenario is related to the way variables are set (Wen et al., 2018).

Despite various usage of techniques and scenarios, one thing can be confirmed is the trend of incorporating accessibility components into hedonic pricing models, where factors affecting housing prices are categorized as three respects: structural characteristics, neighborhood characteristics and locational characteristics. And like what has been done to traditional hedonic models, scholars have improved upon non-spatial regression models in order to catch spatial non-stationary, spatial autocorrelation (SAC) and spatial heterogeneity (SH) of accessibility components that varied over space. Many realized that that the geographically weighted regression (GWR) model is superior, as it focus on revealing local influence of facilities in differentiate locations (Hanink et al. 2012; Nilsson 2014).

This study attempts to take one of the most dynamic cities in China, Shenzhen, as the study object to explore the public service accessibility characteristics of its house price variation, and analyses the association between them. Aim of this study is to monitor the situation of urban public facilities' accessibility impact, achieve better understandings on what drives the house prices in this particular city, and assist in data-driven policy making.

1.2 Scope

This study is scoped to quantitatively analyze the public service facilities accessibility of Shenzhen, and its impact on mechanism at neighborhood level in 2020.

Shenzhen is a major city located in the Greater Bay Area, Guangdong Province, China. It is exceptionally reasonable to take Shenzhen as the research object, as the housing market in Shenzhen expresses significant segmentations and results in uneven distribution of submarket house prices.

The cause of segmentation is initially administerial, tracing back to Shenzhen's being China's first Special Economic Zone (SEZ). A man-made boundary existed between inner-SEZ (327.5 square kilometers) and outer-SEZ (1,692.5 square kilometers). Although this second boundary has been demolished since 2010, its existence still resulted in differential level of economic development, infrastructure condition and social status between inner and outer SEZ. The segregation of public service facilities may or may not be able to be easily eliminated and may results in households' differentiate willingness to pay (WTP), thus making it interesting and meaningful to use urban data and quantitative techniques to verify whether existing literature and views work on the case of Shenzhen.

Besides, Shenzhen is viewed as one of the worlds' most dynamic city, indicating better data situation given at smaller spatial scales. Despite the general data situation in China's cities are bad, lacking a number of data varieties that are commonly found in UK data services, the city's economically activeness still enable this study to combine data from different sources, and partly ensure the credibility of analytical results.

1.3 Research questions and outline

This study hereby aims to address the following questions:

1. What role does public service facilities accessibility plays in a hedonic pricing scenario?
2. What are the spatial effects of public service accessibility on house prices, and what technique is good to address the effects?
3. Which public service accessibility factors have a positive/negative effect on the local house prices in Shenzhen? Do global patterns differ from local patterns?
4. What information is revealed from differentiate neighborhood prices, and how might policy work to make improvements?

The rest of the study is organized as:

Chapter 2: review literature concerning hedonic pricing theory of housing market, its development (improvement of techniques) and its incorporation with public service accessibility.
Chapter 3: describes the case study city Shenzhen and dataset used from three sources.
Chapter 4: introduces the data processing workflow and methodology.
Chapter 5: presents the analysis results of how public service accessibility

characteristics affect house prices in Shenzhen and its influencing mechanism.
Chapter 6: discuss the results and proposed major findings accordingly.
Chapter 7: summarizes the major findings, limitations and future improvements.

Chapter 2. Literature review

2.1 Hedonic pricing theory

Hedonic pricing theory is a well-developed house price modeling technique that can be traced back to Lancaster's consumer theory, in which he argues that utility is generated from properties and characteristic of one certain good (Lancaster, 1966). The hedonic model method started being popular as a tool for property assessment and urban analysis after Rosen (1974) extended it to housing market. As housing characteristics are usually traded in bundles, housing estates are therefore treated as heterogenous goods, in which households aim to maximize the utility through utility-related attributes that form up different prices (Choy et al., 2007). These characteristics mainly consists of a set of structural, geographical, environmental and socio-economic determinants of housing, expressed through the form of an econometric function: A relationship between real estate price P and its associated physical characteristics x_1^P, …,x_n^P as well as neighborhood characteristics x_1^P, …,x_n^P (Helbich et al., 2014). The former depicts property specifics (e.g. floor area, building age, number of stories); the latter refer to the surroundings of properties (e.g. location descriptors, accessibility). For more than half a century, these bundles of attributes have been examined in a number of studies among which these attributes are suggested to be spatially related to location hierarchy.

The spatial immobility of housing implies that location is intrinsic attribute of a house and determines its market values. Bourassa et al. (2003) proposed a notion of substitutability that relates spatial dependence with housing submarkets as a consequence of isolation by spatially differentiate housing locations. Another traditional suggestion made by Straszheim (1975) indicates that housing market are actually clusters of submarkets, differentiated by characteristics of the contained housing units and locations of each submarket. From this view, systematically and geographically varied housing characteristics together with their prices is fundamental to the housing market in an urban area. The research by Grigsby (1963) identifies submarkets according to close substitutability, suggesting that every submarket is a collection of potential dwellings for households to consume with reasonable expectation in a larger market. In particular, Grigsby claims that such collection of dwellings usually fit within a narrow price band. Galster (1996) further developed this view, arguing that the dynamic driving translation of larger housing market into smaller price categories is the 'quality of housing'. He points out that every housing submarket functions as an independent market equilibrating to a specific price dictated by conditions within itself

(Galster, 1996). Thus, one can identify that house prices express spatial variation while occupying complex determinants. Through relationships between house prices and determinants, either positive or negative impacts can be observed and economically valued.

The information collected to characterize housing units possess a strong spatial component, which can be clustered patterns together with heterogeneous structures. Explanation and interpretation of spatial phenomenon for housing will not be straightforward as well, as there is spatial component in house price modelling that requires the use of advanced econometric models. Two current mainstreams to cope with spatial components of house price are: global modelling and local modelling. The two methods each has one certain type of challenge to address. The global method focuses on catching the spatial autocorrelation (SAC) within house prices, which is a concept developed from Tobler's first law of geography (Tobler, 1970: 236): "Everything is related to everything else, but near things are more related than distant things". According to Anselin (1988) and Dubin (1992), SAC describes the coincidence of locational and attribute similarity resulted from similar neighborhood characteristics, socio-economic attributes of households or service quality. In terms of catching the SAC challenge, the traditional ordinary least squares (OLS) method has already been claimed as inefficient and lack of explanatory power (Dubin, 1998). Local modelling methods which focus on analyzing spatially varying relationships, are believed by Fotheringham et al (2002) to have better efficiency to capture spatial effects. In classic hedonic house price research, most hedonic models are used to measure global relations between house prices and indicators, while more and more literature started fitting hedonic house price models with local econometric models

2.2 Modeling spatial effects in Hedonic models

Hedonic pricing models are used to be expressed as linear regression models, in which house prices are regressed on housing attributes including structural, socio-economic and surroundings characteristics. Within a traditional hedonic model, model calibration mainly adopt the ordinary least square (OLS) method, in which estimates of a parameter is explained as a portion of a house's overall price given by certain characteristics (Hulten, 2003). It can be also explained as buyers' willingness-to-pay (WTP) for an additional unit of specific characteristic. Considering the spatial component of housing market, spatial econometric methods should have better performance than OLS. Unlike traditional nonspatial econometric modelling, spatial regression modelling try to cope with two intrinsic spatial effects: spatial autocorrelation (SAC) and spatial heterogeneity (SH) (Anselin, 1988). SAC recognizes the fact that one determinant measured at a given location is spatially correlated with the same determinent measured nearby. In turn, SH indicates relationships (expressed by parameter estimates in regression models) that change with space.

The geographically weighted regression (GWR) is a local modeling approach which allows variation of parameter estimates over space (Brunsdon et al. 1996; Fotheringham and Brunsdon 1999; Fotheringham et al. 2002). Instead of using one single model to describe the entire housing market, the GWR approach create a separate model for every single spatial unit, and weights other observations according to their distance to that unit. The GWR approach uses housing parcel centroid coordinates, which makes it a special case of locally weighted regression that are applied to geographic space (McMillen and Redfearn, 2010). The approach is non-parametric, which means it does not require any assumptions to be made (which are necessary for OLS models) considering the underlying distribution of predictors values. It also has the ability to handle highly skewed or categorical predictor variables. By using GWR as an exploratory tool, researchers are able to understand the variation of preferences for specific property attributes; and by utilizing it statistical power, GWR is able to support estimation of house price with given sets of attributes with location effects taken into account.

The superiority of GWR in catching spatial effects will provide urban planners with valuable information, while in a context of housing market, the generated information could be residents' willingness to pay (WTP). In an example discussed by Crespo and Grêt-Regamey (2013), citizens have different preferences on populated areas, resulting in up and down, positive and negative local GWR estimates that vary over space; Conversely, for population don not want wish for proximity to certain places, the GWR estimates may also express discrepancies.

Although GWR has already shown superiority, it is still necessary to push forward wider use of this method. Bitter et al., (2007) argues that future studies should provide direct comparison between GWR and outdated methods, while McCord et al. (2012) calls for more examination to evaluate the GWR method. With rich literature in hedonic pricing theory and justified technique of GWR, it is reasonable to expect more usage and studies regarding house price modelling to come out.

2.3 Modeling accessibility in hedonic house price

Classical Ricardian value theory which developed from neoclassical microeconomic theory (Alonso, 1964), has long emphasized the 'access-space’ trade-off as one fundamental urban law. In this equation, higher transport costs balance lower land costs in less accessible parts of a city. When fitting this insight into housing market research, it actually says land and accessibility are substitute hedonic factor in shaping house prices. However, from today’s view Alonso’s (1964) monocentric model does not adequately capture accessibility value as workplaces are not solely located in city center, or any specific point (McDonald, 1987), and so do residences. The multicentric nature of housing-market areas increased the dimensions that are required to be taken into account when studying accessibility factors, meanwhile motivated the search for alternative methods that go beyond measuring physical distance and travel time with distinctly defined centers.

As more and more studies during the development of hedonic pricing theory reveal that housing can not be regarded as a homogeneous good, but one comprising various characteristics, the concept of accessibility can also find its place as utility attributes (Rosen, 1974) valued through hedonic pricing. An urban area can be treated as one single market, and then broken into smaller parts, relating to attributes of neighborhood conditions, public service utilities and accessibility. House purchasers may increase their preference on any of the characteristics by choosing from a range of alternative locations or properties. In such an action, distance or accessibility is regarded as an influence, and perhaps a major influence considering its critical role to land values in the trade-off between travel costs and housing costs (Alonso, 1964; Muth, 1969). Moreover, Beckmann and Martin (1973) made a central assumption that housing and accessibility are jointly purchased in a residential choice decision. Michelson (1977) in his discussion of individual evaluation on residential purchase, found that quality of housing mattered more for suburban dwellers while distance took priority for those in the city, indicating valuation of accessibility varied over space.

Accessibility-space literature has been linked to hedonic modelling in a more recent study by Evans (1995), in which he argues that hedonic models developed the trade-off models from simply location-based to a range of location and dwelling characteristics with extended determinants of price. Meanwhile, much of the data used in hedonic analysis lacked land and location information (Cheshire and Sheppard, 1997), calling for appropriate incorporation of accessibility information. Other than the traditional measure of distance to CBD which is just a simple approximation of travel cost, Gibbons and Machin (2005) explored the effects of accessibility on house prices and proposed two ways to define access: distance to the target and service frequency at the nearest targets. Such definition not only is an evolved method for studying travel accessibility, but also shine the path to explore further external effects such as being accessible to urban public services, and their impact on house prices.

Urban public service facilities are basic factors affecting the living quality of residents and indices of attractiveness for different locations. A wide range of studies have been carried out towards the external effects brought by urban public service resources towards housing prices. Focuses of these studies has been education, green space, the underground and various public service facilities that are closely related to citizens' daily activities, many found that accessibility towards certain public service (e.g. schools and education) positively impact housing prices (Lan et al., 2018). Whereas in these studies, two main methods are used to quantify the accessibility of a service: one is using a combination of distance, commuting time or expense of a mean of transport to represent accessibility (Mok and Ling-Hin, 2010), another is measuring accessibility by the availability or proximity of relative resources within a given distance (Wen et al., 2018). There are also cases that certain service does not significantly affect housing price indicating that every time the scenario is related to the way variables are set (Wen et al., 2018).

One thing can be confirmed is the trend of incorporating accessibility components into hedonic pricing models, where factors affecting housing prices are usually categorized as: structural characteristics, neighborhood characteristics and locational characteristics. Previous works suggest public service facilities used to be included in the model as neighborhood characteristics (Lin et al., 2013; Melichar, 2013). And like what has been done to traditional hedonic models, scholars have improved upon non-spatial regression models in order to catch SAC and SH within accessibility components that varied over space. Many realized that the geographically weighted regression (GWR) model is superior, as it reveals local influence of facilities at each specific location unit (Hanink et al. 2012; Nilsson 2014).

2.4 Conclusion

Consider the rich literature in hedonic pricing theory, and growing usage of the GWR techniques, it has been clear that a combination of Hedonic housing characteristics and GWR model could be the suitable way to study house price variation in cities. Meanwhile hedonic models are being used frequently as well in accessibility analysis (Adair et al., 2000). The advantage of applying hedonic modelling in accessibility studies, is that it helps urban configuration of the study area, which is previously difficult (Xiao et al., 2016). By using GWR to study accessibility within hedonic models, the locational externalities can be better specified, with spatial effects taken into account leading to less biased results.

Chapter 3. Study Area and Data

3.1 The Shenzhen housing market

Shenzhen is a major city located in the Greater Bay Area, Guangdong Province, China. The city situates in the central coastal area of southern China, neighboring Hong Kong, with a land area of 2,050 square kilometers (Zhou et al., 2010) including totally 10 districts: Luohu, Futian, Nanshan, Yantian, Longgang, Bao'an, Longhua, Guangming, Pingshan and Dapeng (Figure 1). It is an overall small-sized metropolitan. Formation of the city is the result of a rapid state-driven urbanization since the mid-1980s, with increasing economic development and receiving continuous attention.

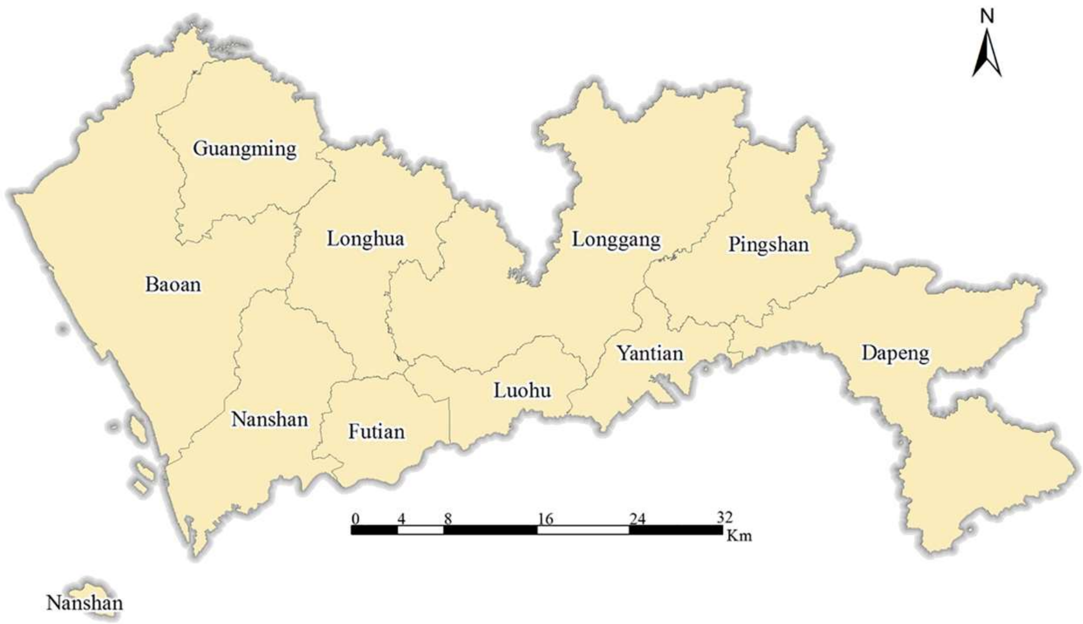

Figure 1. Administrative districts of the case study city Shenzhen

Highly populated areas do not distribute evenly for Shenzhen, as most of the population concentrate in southern part of the city, resulting in densely situated neighborhoods. One main reason for formation of such distribution can be traced back to the founding of China's first Special Economic Zone (SEZ), in which Shenzhen has been transformed form an agriculture-based rural area into a metropolis under planned investments (Xu, 2008). Before the year 2010, the nature of SEZ created a physical boundary to identify the area where special economic activities and social policies were applicable. Therefore, in addition to the city's administrative border with other regions, Shenzhen has a second border inside the city which divided the whole city into two parts: the SEZ (Futian, Nanshan, Luohu and Yantian districts, 327.5 square kilometers) and outer-SEZ (Longgang, Bao'an and remaining districts, 1,692.5 square kilometers). Although this second boundary has been demolished since 2010, its existence still resulted in differential level of economic development, infrastructure condition and social status between inner and outer SEZ.

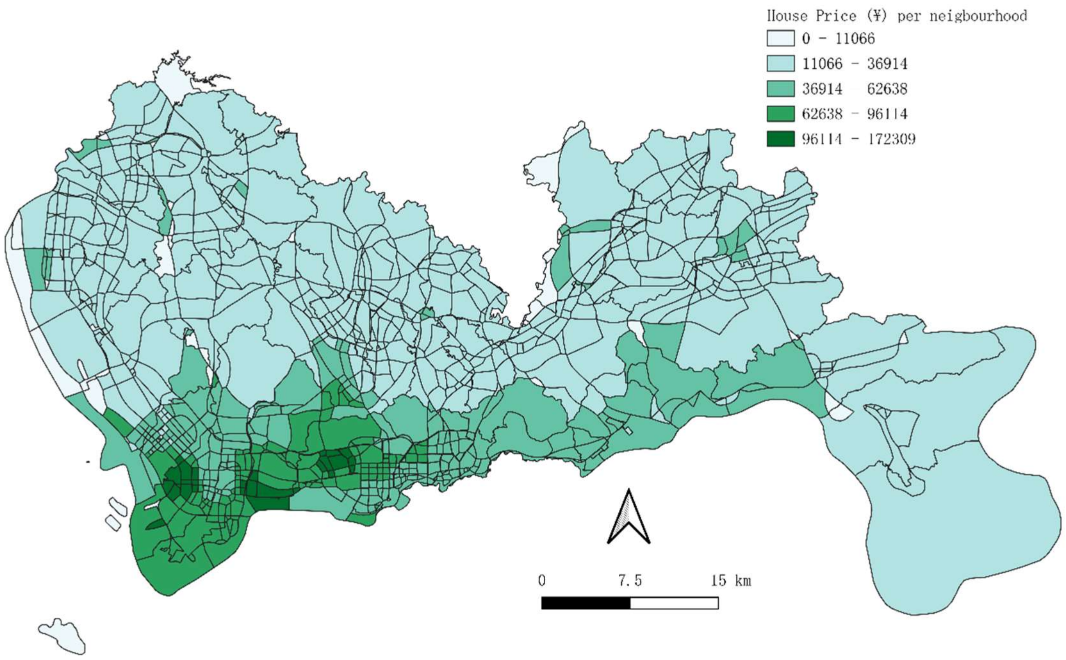

Figure 2. Shenzhen house price distribution at neighborhood level

Influenced by the inner-outer SEZ segmentation, the housing market of Shenzhen exhibits distinctly divided submarkets (Figure 2). The most expensive housing mainly situated in inner areas (Futian, Luohu and Nanshan) of Shenzhen, expressed as higher prices for every floor area unit (square meter). Causes of house price segmentation are partly historical, as in Chinese cities, residential buildings constructed before the 1990s housing reform were mainly old blocks (Gaubatz, 1999). The old housing were used to be assigned to workers according to individual service duration as a component of state welfare (Zhang, 2000), indicating little attention paid to the location and quality of housing (Jim and Chen, 2007). After the housing marketization reforms in 1990s when housing and land markets were introduced, commodity housing has emerged and households are encouraged to satisfy their specific housing demands, meanwhile had to endure the fact that newly-built housing with less service density generally locate outwards from the inner city.

As one of the worlds' most dynamic city (Wu et al., 2016), together with large amount of young population, Shenzhen offers itself as a good sample for investigating the influence of urban service hot spots on house prices. The levels of POI activity change dramatically with housing prices increasing (Wu et al., 2016). Therefore, this study chooses Shenzhen as the sample city because the relevant data for studying public service facilities are more updated and meaningful compared to that in less economically active cities.

3.2 Data selection and pre-processing

3.2.1 The Shenzhen Road network map

The Road network map of Shenzhen, which contains 1184 polygons, is obtained from the Urban spatial informatics Lab of Shenzhen University. The R3 road network map is the second smallest spatial unit available for analyzing spatial effects in the Greater Bay Area (GBA). Since there is no neighborhood tabulation map or similar alternatives, this study uses the road network map (Figure 3) as it is capable of representing the rough coverage of a neighborhood and in the meantime, avoid areas that are not suitable for human habitation. This study treats every polygon of the R3 road network map of Shenzhen as a neighborhood, with an average of 0.724 square kilometer per polygon. Each unit is labeled with unique parcel ID by the data provider, indicating convenience to merged and used in different projects or switch to larger spatial units. The original coordinate reference system of the map is WGS 84(ESPG: 4326), and has been transferred into projected coordinate system WGS84/UTM Zone 50, in order to enable further analysis regarding distance calculation.

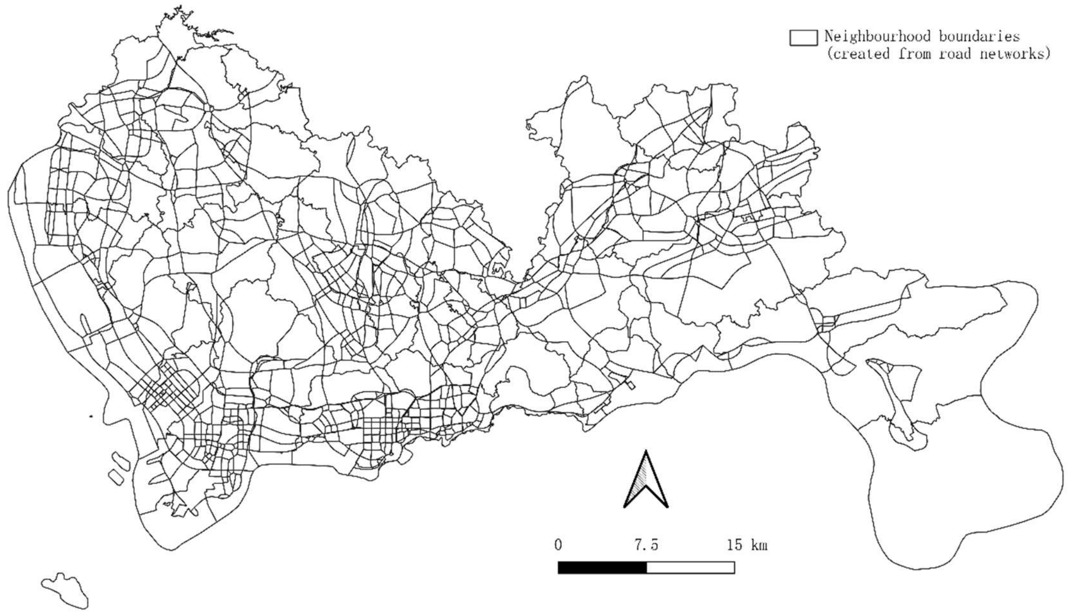

Figure 3.Neighborhood boundaries (defined by road networks)

3.2.2 The Anjuke house price data for residential communities

The official statistical system in China has not been well developed and there is no census data of house price that can be directly used to represent the house price at neighborhood level. In this case, this study has to carry out some data pre-processing to transform house price data of different forms into neighborhood level house prices that can be utilized to understand spatial effects among neighborhoods.

Anjuke[1] is a leading domestic real-estate information service platform in China, founded in 2007. The Anjuke website has done massive data collection on building information and built up authoritative real-time databases for properties in 640 cities of China. This study utilizes the information on the 'residential community navigation' page of Anjuke Shenzhen website, by scraping relevant information from the site's HTML document using PyCharm. The scraped information includes 8 fields (Table 1): District name, Sub-district name, Name of Community, Year of completion, Address, Price, Longitude and Latitude, all of which are included in the navigation page. Information of 8750 residential communities was scraped from the web and exported into a Comma-Separated Values (CSV) file for further processing and analysis.

[1] https://shenzhen.anjuke.com/community/?from=navigation

Table 1. Anjuke house price data scraped from web

Column	Description	Year/Month
DS	Administrative district of the residence	2021/6
JD	Sub-district of the residence (Jiedao)	2021/6
name	Name of the residential community	2021/6
finish_date	The year of completion of that residence	2021/6
address	Disrict-Subdisrict-Road	2021/6
price	Chinese Yuan (¥) per square meter	2021/6
latitude	Latitude of the residence	2021/6
longitude	Longitude of the residence	2021/6

3.2.3 The Baidu Map Points of Interest (POI) data

The Baidu Map[2] is a desktop or mobile web mapping service application and set of technologies, provided by Baidu. The application offers whole sets of services including satellite imagery, street maps, street/indoor view and route planner. The main focus of the App had been mainland China, Hong Kong, Macau and Taiwan before 2016, while now it supports mapping of over 150 countries. The Baidu Map also utilizes map data supplied by multiple sources (e.g. OpenStreetMap).

Points of Interest data, namely POI data, can be one of the followings: one single building; a tourist spot; a mailbox or a bus station.
This study utilizes the Baidu Map POI retrieval tool, and retrieved POI data according to the default categories set up by the platform. The retrieved data are included in Table 2.

[2] https://lbsyun.baidu.com/index.php?title=androidsdk/guide/search/poi

Table 2. Retrieved data from Baidu Map platform

POI name	Description	Year/Month
EDU	Public schools; colleges; education agencies…	2020/7
FOOD	Restaurants, food stores, snack shops…	2020/7
SHOP	Convenience stores, supermarkets, outlet stores…	2020/7
LEIS	Internet cafes, bars, recreation clubs…	2020/7
BANK	Community financial facilities	2020/7
MALL	Shopping malls, hypermarkets	2020/7
BLD	High-rise buildings, skyscraper	2020/7
GOV	Governmental buildings	2020/7
TOUR	Tourist spots	2020/7
METRO	Metro stations	2020/7
BUS	Bus stations	2020/7

Chapter 4 Methodology

4.1 Selecting variables for accessibility measurement

The Standard for Planning and Design on Urban Residential Area (2018) released by the state government of China, has specified a service radius for public service facilities considering a 15-minute walking distance. That document certify that public service facilities should be within a resident's 15-minute walking distance. That regulation gives this study a criterion to measure accessibility of public service facilities by calculating the quantity of facilities with a 15-minute walking range. The measurement of radius is based on the suggestion that with a walking speed of 4km/h, a household member is able to travel 1km within 15 minutes (Wen et al., 2018). Thus, this study uses the number of facilities (Baidu POIs) within 1km around a residential community (Anjuke residential points) to quantify the accessibility of public service facilities. This study merges the calculated buffer counts into the neighborhood map (road network map). The calculation is realized through the buffer analysis tool in QGIS. By calculating the number of specific POIs within 8750 buffer areas and aggregating them into 1184 neighborhoods (Figure 4), 11 variables are created representing the accessibility to certain public service facilities for each neighborhood.

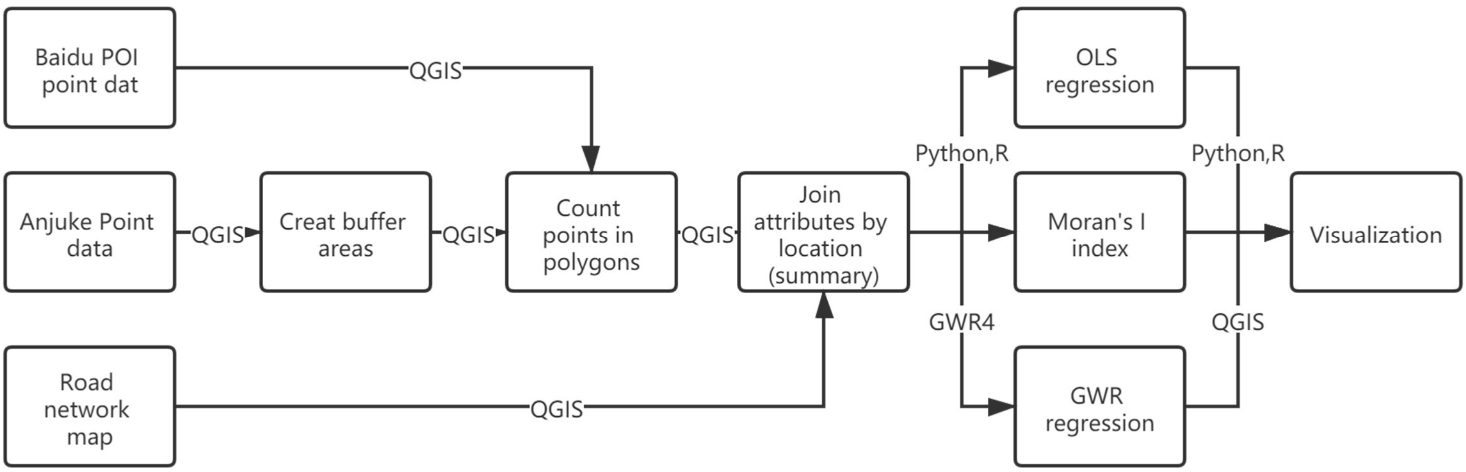

Figure 4. Data processing workflow

Following the hedonic pricing theory, totally 14 variables (Table 3) concerning structural attributes, neighborhood attributes and locational attributes are selected as explanatory variables. Among the three sort of indicators, the accessibility of 11 kinds of public service facilities is defined as 11 neighborhood characteristic variables, which will support the main focus of this study. The other 3 variables (1 structural and 2 locational) will be used as control variables. The ‘Age’ variable is calculated from the ‘finish date’ column in Anjuke scraped data, indicating the number of years a residence has stand from its completion. The ‘DistCBD’ variable is calculated by applying a distance matrix to every neighborhood to measure their distance to centroids of Futian, Luohu and Nanshan districts. The ‘DistGreen’ variable is calculated from applying a distance matrix to every neighborhood to measure their distance to the nearest green land polygons.

Table 3. Summary statistics of selected variables (after processing)

Characteristic	Predictors	count	mean	std	min	25%	50%	75%	max
Structural	age	1184	7.29	5.56	-2	2.60	5.95	12.44	27
Neighbourhood	EDU	1184	29.70	30.08	0	6	19.17	46.53	140.14
	FOOD	1184	124.47	130.06	0	30.38	79.50	165.29	731.77
	SHOP	1184	335.29	355.62	0	74.18	201.50	474.73	1624.56
	LEIS	1184	86.21	87.83	0	22.51	56.20	118.57	555.86
	BANK	1184	32.44	35.21	0	8.77	20.77	42.17	194.24
	MALL	1184	20.64	16.17	0	7.90	17.85	30.23	94.56
	BLD	1184	16.33	23.37	0	1.50	6.99	22.80	129.74
	GOV	1184	11.82	16.25	0	1.50	4.59	13.21	103.00
	TOUR	1184	1.38	7.21	0	0	0.25	0.73	105.00
	METRO	1184	3.33	3.54	0	0.65	2.14	4.63	16.61
	BUS	1184	212.04	132.42	0	110.78	191.02	312.52	586.77
Locational	DistCBD	1184	11718	9382	0	3706	8794	19903	37744
	DistGREEN	1184	840	599	0	407	693	1170	4358
Response variable	Ln(price)	1184	10.00	2.29	0	10.1	10.36	10.91	12.06

4.2 Investigating spatial relationship

4.2.1 The OLS regression

Firstly, the Ordinary Least Squares (OLS) regression is used to investigate the global relationship between house price and accessibility factors. The OLS regression examines the average state of parameter estimates. It can be written as:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_i x_{ik} + \varepsilon \qquad i = 1,2,3, \ldots, n \qquad (1)$$

Where β_0 indicates the intercept, β_i indicates the ith estimated parameters, x_{ik} is the kth accessibility factors of the ith neighborhood, ε indicates the error term.

In addition to the equation, the value of average house price per every neighborhood is log-transformed in order to normalize its positively skewed distribution, and is used as the output variable y_i. For neighbourhoods missing house price data before and after the logarithmic transformation, zero values are kept or added to the original data.

4.2.2 Testing Spatial Autocorrelation with Global Moran’s I

The assumption of Independence of Errors states that residual values (errors) in the OLS model must not be correlated in any way. If they are, however, then there may exists autocorrelation, indicating there are factors going behind the scenes that have not been sufficiently accounted for in the model. Since the data used in this study are spatially referenced, and a large number of studies regarding hedonic house price have already mentioned the intrinsic spatial autocorrelation (SAC) in the spatial distribution of house price and indicators. Therefore, spatial correlation test is required to check whether there exists such spatial effect, and to what extents it may affect the regression model. It is commented that in the case of SAC exists, the Geographically weighted regression (GWR) method may be more appropriate than the OLS regression as GWR account for the information on spatial variation.

Global Moran’s I index (MI) is used in this study as it measures whether the house prices of Shenzhen at the neighborhood level are spatially correlated, it can be expressed as:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \qquad (2)$$

In which:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}) \qquad (3)$$

and

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad (4)$$

Where n indicates the number of neighborhoods, Y_i indicates the residuals in the OLS regression, w_{ij} is a spatial weight matrix. The Criteria for interpreting Moran's I index are shown in Table 4.

Table 4. Criteria for interpreting Moran's I index

Value	Correlation pattern	Significance interval
0<MI<1	positively autocorrelated(clustered)	p<0.05
MI=0	distribute randomly	
-1<MI<0	negatively autocorrelated(dispersed)	

This study uses a k-nearest neighbors of 4. The neighbors are measured by the distance between centroids of each area through RStudio. The spatial weight matrix to the neighbors is constructed by adopting row-standardization to calculate the Moran's I index.

4.2.3 Geographically Weighted Regression (GWR)

1. Base model

Different to OLS regression, the GWR considers the spatial location of observations and performs independent linear regression for each observation with unique equations for each. The GWR model output a set of coefficient estimates showing how the relationship between input variables and response variable varies over space, thus is able to explain the differentiation in each variable's influence among regions. The GWR model is given as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \qquad i = 1,2,3, \dots, n \qquad (5)$$

Where (u_i, v_i) is the coordinate of the ith neighbourhood, $\beta_0(u_i, v_i)$ is the intercept for neighbourhood i , $\beta_k(u_i, v_i)$ indicates the local parameter estimate for explanatory variable x_k at neighborhood i , ε_i indicates the ith error term. In addition to the equation, the value of average house price per every neighborhood is log-transformed in order to normalize its positively skewed distribution ( ), and is used as the output variable y_i . For neighborhoods missing house price data before and after the logarithmic transformation, zero values are kept or added to the original data.

2. Selection of Kernel function and bandwidth

(1) Kernel function
Spatial weight matrix is a core component of a GWR model, calculated from specific kernel function. Different kernel functions imply different understandings of the spatial relationship within data, resulting in different spatial weight matrix. The Gaussian Function is one of the most popular selections among researches, it can be written as:

$$w_{ij} = exp^{-(\frac{d_{ij}}{b})^2} \qquad (6)$$

Where w_{ij} indicates the weight matrix between observations i and j , d_{ij} indicates the Euclidean distance between i and j , and b indicates the bandwidth.
Two kernel types are selectable for the kernel function within the tool GWR4: fixed bandwidth and adaptive bandwidth, indicating a fixed bandwidth size defined by a distance metric measure, or an adaptive bandwidth size defined as the distance to the kth nearest neighbor to calculate equation (6).
After having picked the bandwidth, larger d_{ij} will lead to smaller weight between observations i and j . Distance between two observations i and j exceeding the bandwidth threshold will result in their weight being close to 0. In the case of urban public service accessibility, the accessibility for a neighborhood is affected by neighboring neighborhood's accessibility, not likely to exceed the threshold.

(2) Bandwidth criteria
GWR is particularly sensitive to the bandwidth. Too big bandwidth result in excessive deviation of parameter estimate, while too little bandwidth result in excessive variance of parameter estimation. Thus, selection of bandwidth is of great importance. Two common selection criteria are the Cross Validation (CV) and Akaike Information Criterion (AIC). Whereas for the AIC criteria, there is an option of AICc (small sample bias corrected AIC) which is more suitable for predicting local Gaussian regression modelling, as classic AIC tends to choose smaller bandwidths by which geographically varying

coefficients are likely to be undersmoothed. For the same data sample, the optimal choice of bandwidth will be the one that minimizes the value of CV or AIC.

In this study, both CV and AICc optimization are used to find the optimal fixed or adaptive bandwidth with the Gaussian kernel function. The bandwidth that brings the GWR model with largest R-square value and smallest CV or AICc will be selected for parameter estimation.

It is worth mentioning that the OLS regression still gives information even the study uses GWR regression. The OLS results reflect the impact of public service accessibility on neighborhoods' house price at a global level, and help pick the explanatory variables through checking multicollinearity. Additionally, by comparing the results of GWR with the results of the OLS model, the unique role of the GWR model is better demonstrated, through outlining its advantages in explaining spatial differentiation and improving interpretation effect.

Chapter 5 Results

5.1 Correlation matrix and Variance inflation factor (VIF) test

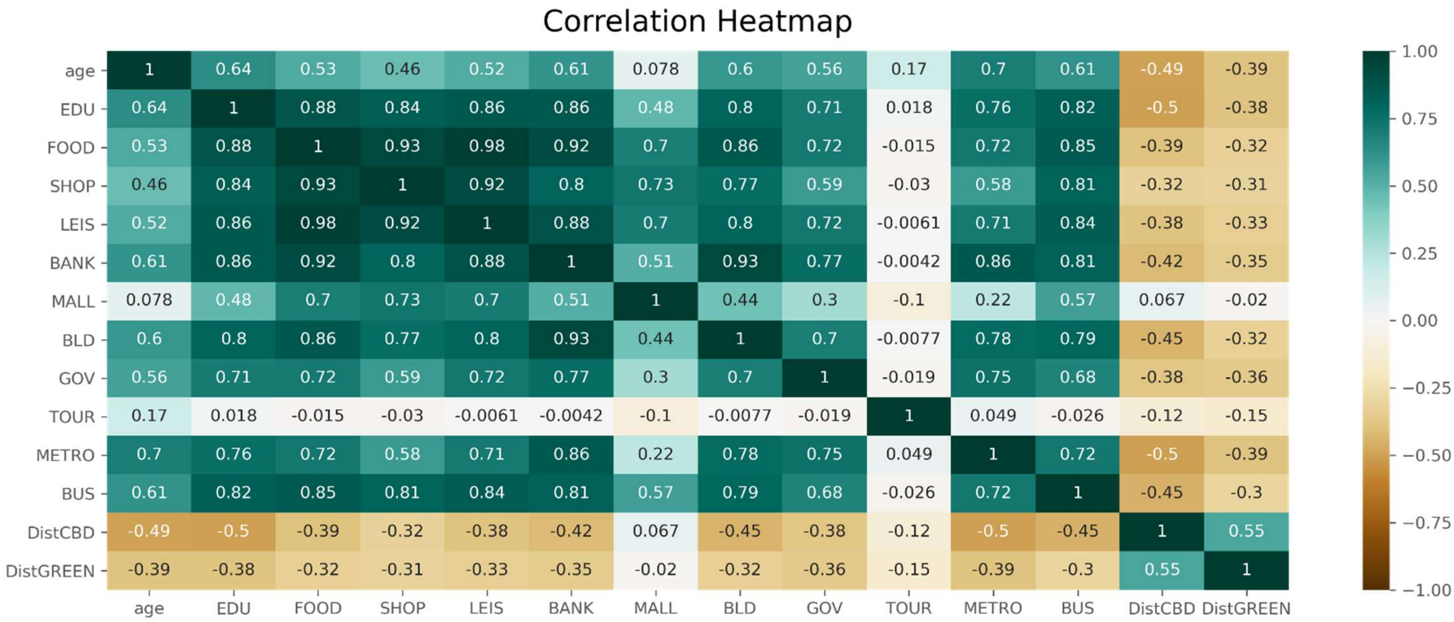

Figure 5. The correlation matrix of structural, neighborhood and locational characteristics

Table 5. Results of VIF test after 4 removals

Characteristic	Predictors	VIFs	VIFs	VIFs	VIFs	VIFs
Structural	age	2.5082	2.4955	2.4807	2.4758	2.4751
Neighborhood	EDU	7.2188	6.8311	6.4254	6.3755	4.7742
	FOOD	63.7316	/	/	/	/
	SHOP	12.6342	11.4493	10.8370	7.7039	/
	LEIS	36.0564	17.0545	14.3183	/	/
	BANK	24.7010	22.8049	/	/	/
	MALL	4.6415	4.4007	4.3222	3.4285	2.3638
	BLD	11.4706	10.4884	4.4602	4.4599	3.9491
	GOV	3.1519	3.1516	3.1477	2.6683	2.6597
	TOUR	1.0901	1.0898	1.0893	1.0820	1.0820

	METRO	6.5491	6.4442	4.8763	4.3679	4.2369
	BUS	5.1853	5.1305	5.0674	5.0669	4.9669
Locational	DistCBD	2.3061	2.2579	2.1607	2.0804	2.0763
	DistGREEN	1.6088	1.5881	1.5776	1.5756	1.5292

Figure 5 shows the results of Pearson's correlation calculated from Python. The results shows that the number food, shop, leisure and finance facilities within a 1km neighborhood radius is positively correlated to all other neighborhood characteristics. The number of tourism facilities within a neighborhood of 1km have very weak correlations with all other attributes. The two locational characteristics: Distance to CBD and distance to green space have negative correlations with building age and all facilities attributes, indicating service frequency is likely to increase when approaching CBDs and green space.

By checking the multicollinearity between all predictors using a VIF test, it is observed that food, bank, leisure and shop facilities contribute greatly to the multicollinearity between the predictors consider a threshold of 5 (Table 5). After removing these 4 predictors, the final VIF test shows that all remaining predictors do not exceed the threshold of 5, thus, are kept for further analysis.

5.2 OLS Regression results

Table 6. The OLS regression results

Characteristic	Predictors	coefficient	Std.error	t	P> | t |
	Constant	4.5144	0.156	28.991	0.000
Structural	age	0.1789	0.013	14.139	0.000
Neighbourhood	EDU	0.0098	0.003	3.032	0.002
	MALL	-0.0074	0.004	-1.734	0.083
	BLD	-0.0338	0.004	-8.901	0.000
	GOV	-0.0128	0.004	-2.843	0.005
	TOUR	0.0365	0.006	5.666	0.000
	METRO	0.1251	0.026	4.808	0.000
	BUS	0.0099	0.001	13.207	0.000
Locational	DistCBD	9.082e-05	6.86e-06	14.282	0.000
	DistGREEN	0.0012	9.23e-05	13.212	0.000

R-squared	Adj.R-squared	F-statistic	Prob(F-statistic)	AIC
0.553	0.549	145.0	4.83e-197	4389

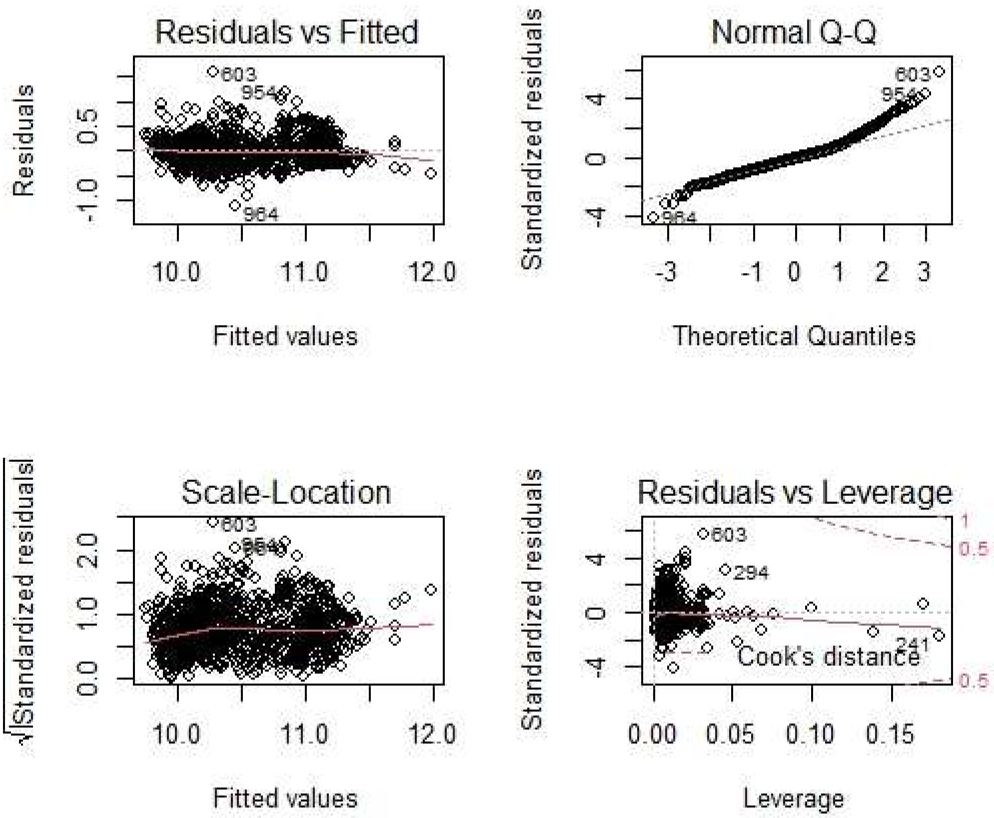

Figure 6. Residuals vs. fitted ln(price)

Table 6 shows the results of the OLS regression of 10 dependent variables. The R-squared value of 0.553 indicates the model is able to explain variation of public service facilities accessibility in Shenzhen's neighborhood at 55.3%. The p-values of all predictors are smaller than 0.05 except for mall facilities, whose p-value is smaller than 0.1. Therefore, it is very likely that the observed relationship is not random, and the proposed model is overall significant. The Figure 6 shows that the model does not violate regression assumptions.

The results show that building age has globally positive influence upon house prices. As a city with 40 years urbanization history, it might be the case that older buildings mainly concentrate at the inner city where service frequency is better than outer regions, thus leading to higher prices. The development and massive residential housing construction of the outer city (outer-SEZ) is usually 20-30 years later than inner-SEZ, thus may be the leading to this global trend of older building being more expensive.

The results show that number of education facility, tourist facility, metro station and bus station also have positive influence upon house prices. That is, higher service frequency of these facilities within an accessible radius of 1km lead to higher house prices. Basically, such results conform to Shenzhen’s scarce educational resources and high demand for public transport.
It is observed that higher service frequency of shopping malls, high-rise buildings and governmental buildings within an accessible radius of 1km lead to lower house prices. Such results violate several studies that were carried out towards the same kinds of public services regarding sample cities in China. Densely located shopping malls, high-rise buildings and governmental buildings usually undertake higher land prices that should have driven house prices higher.

Additionally, the results show that distance to CBDs and green space have weak positive influence on house prices. According to previous figure 1, it was observed that residential communities with higher house prices obey the inner-outer city segregation and mainly concentrate in the CBD districts (Futian, Nanshan, Luohu), indicating the global regression results fail to address minor segmentations of Shenzhen’s housing market.
Overall, the results of OLS regression explain some of the variation but fail to give reasonable explanations concerning every predictor. That motivate the study to carry out more advance techniques to address local spatial differentiation and spatial effects in order to give better explanations.

5.3 Global Moran's I and GWR results

5.3.1 Global Moran' I Index

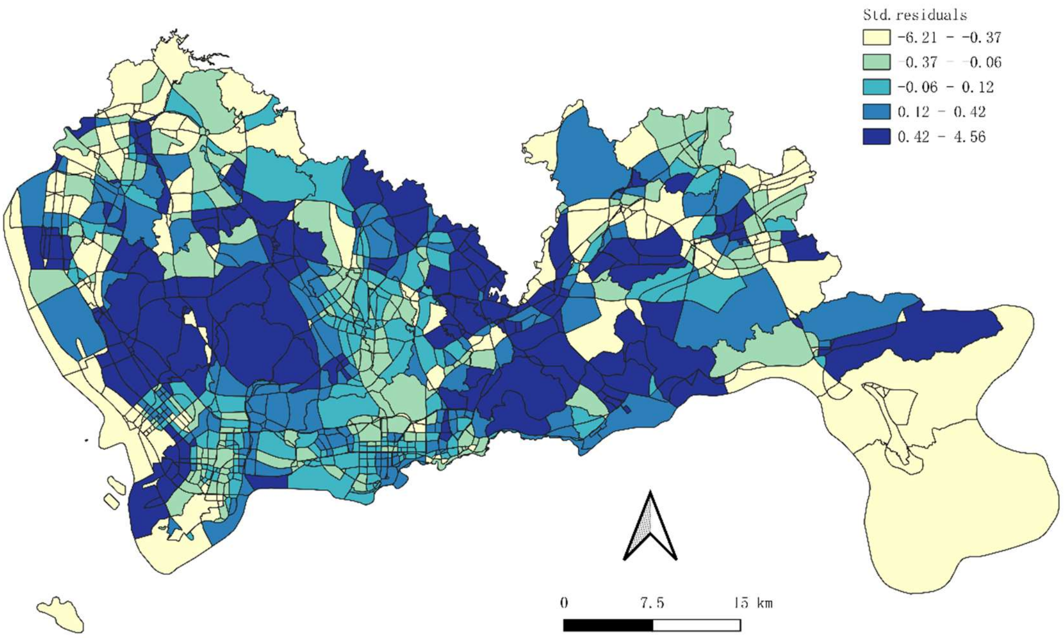

Figure 7. Standardized residuals of the OLS model

The standardized residuals map of the OLS model (Figure 7) is plotted. The residual map shows that residuals do not distribute randomly in Shenzhen neighborhoods. The Moran's I index shows that there is a weak to moderate (MI=0.485, $p<0.05$) positive spatial autocorrelation exists for the residuals. This suggests that the global-level OLS regression model is biased, indicating improvements to be made on catching spatial variation and effects by employing a GWR model.

5.3.2 GWR Regression

1.Bandwidth selection

Table 7. Selections of model parameters

Criteria	No.	Kernel type	Bandwidth size	Residual squares sum	Sigma	AIC	R-squared
CV	1	Fixed Gaussian	2692.58	97.14	0.29	859.95	0.98
	2	Adaptive Gaussian	62.82	878.07	0.86	3171.00	0.86
AICc	3	Fixed Gaussian	2692.58	97.14	0.29	859.95	0.98
	4	Adaptive Gaussian	62.82	878.07	0.86	3171.00	0.86

Table 7 shows the goodness of fit with different parameters. It can be observed that model 1 and model 3 has the largest values of R-squared and smallest AICs. At the same time, their residual sum of squares and sigma estimates are the smallest. According to this result, this study adopts the optimal bandwidth (GWR4 golden selection) obtained from the AICc criteria and the Gaussian kernel function with a fixed bandwidth size to estimate the GWR model.

2. GWR results

Table 8. Quantile statistics of regression coefficients in GWR model

Characteristic	Predictor	Min	25%	50%	75%	Max
	Intercept	0.0000	0.8757	5.2140	9.6249	12.0910
Structural	age	-0.2025	-0.0139	0.0281	0.1426	0.6620
Neighbourhood	EDU	-0.7204	-0.0089	0.0071	0.0253	0.5703
	MALL	-0.2239	-0.0258	-0.0080	0.0038	0.4848
	BLD	-1.4355	-0.0193	-0.0034	0.0253	0.8707
	GOV	-0.5920	-0.0412	-0.0091	0.0076	0.2327
	TOUR	-3.4147	-0.0310	0.0133	0.3090	3.7567
	METRO	-1.4482	-0.0374	0.0086	0.1753	1.1203
	BUS	-0.0344	0.0006	0.0041	0.0101	0.0441
Locational	DistCBD	-0.0002	-0.00004	0.0002	0.0003	0.0014
	DistGREEN	-0.0012	0.0001	0.0005	0.0012	0.0052

From Table 8, we can see that the coefficients of predictors vary greatly, with all predictors having both negative and positive coefficients. The results show that in spatially different neighborhoods each predictor has different degrees and trends of influence. The structural, neighborhood and locational characteristics all present significant variation indicating there is spatial heterogeneity within the original data.

Table 9. Comparison of results goodness of fit between GWR and OLS models

	GWR	OLS
R-squared	0.98	0.55
AIC	859.95	4389.00

Table 9 shows that the R-squared value of the GWR regression is higher than that of OLS model, while the AIC of GWR model is smaller. This indicates that the GWR model performs better than the OLS model in both goodness of fit and interpretation effects.

3.Spatial distribution of the coefficients

(1) Accessibility to education facilities

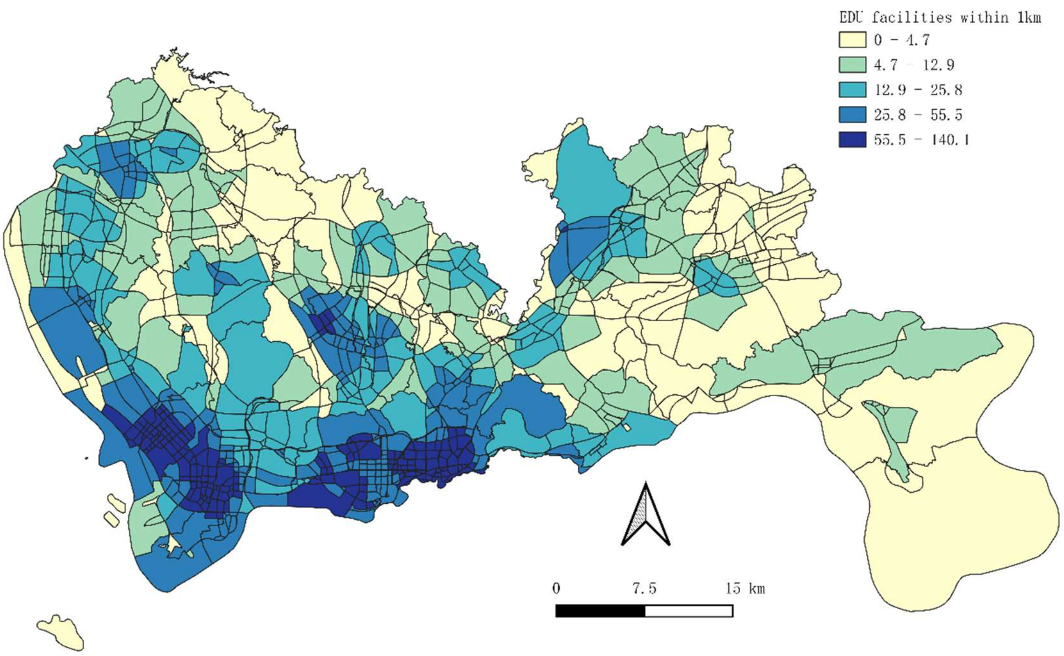

Figure 8. Average number of education facilities within 1km radius of each neighborhood

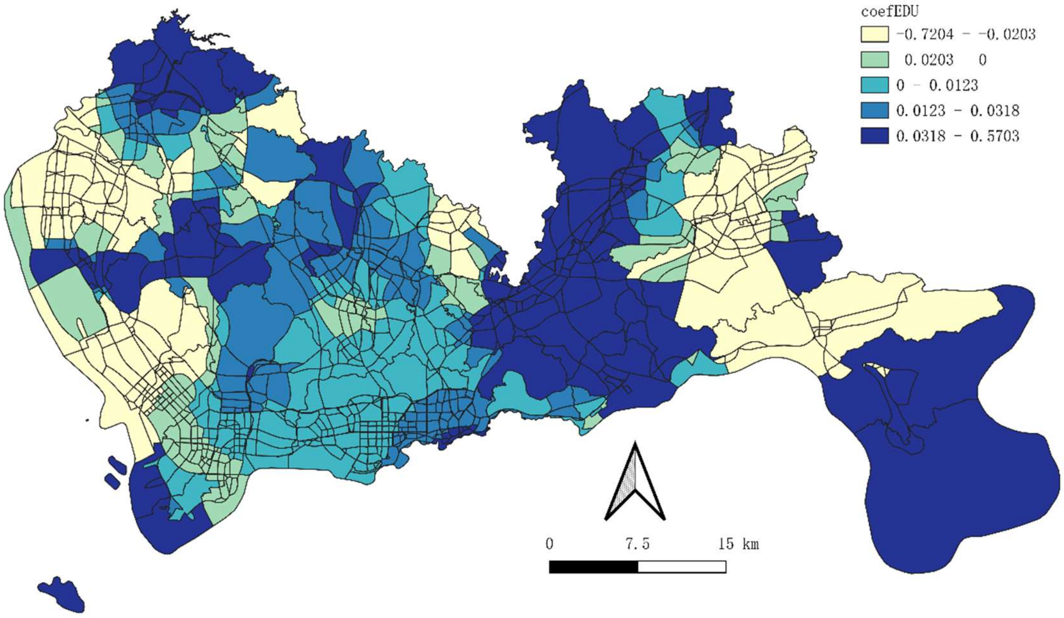

Figure 9. Distribution of coefficients (effects of education facilities on house prices)

It can be observed from Figure 8 and 9 that education facilities mainly aggregate at inner-SEZ districts, while for northern, north-east and eastern districts, accessible education facilities for local neighborhoods are few.
The corresponding coefficients distribution show that education facilities accessibility has greater influence on house prices in these less accessible districts. As for inner-SEZ districts and most other neighborhoods, proximity to education facilities still matters a lot and have positive influence on house price. For some neighborhoods in Futian and Luohu districts, even they already are highly accessible to education facilities, their house prices are still heavily positively influenced by such an advantage.

(2) Accessibility to shopping malls

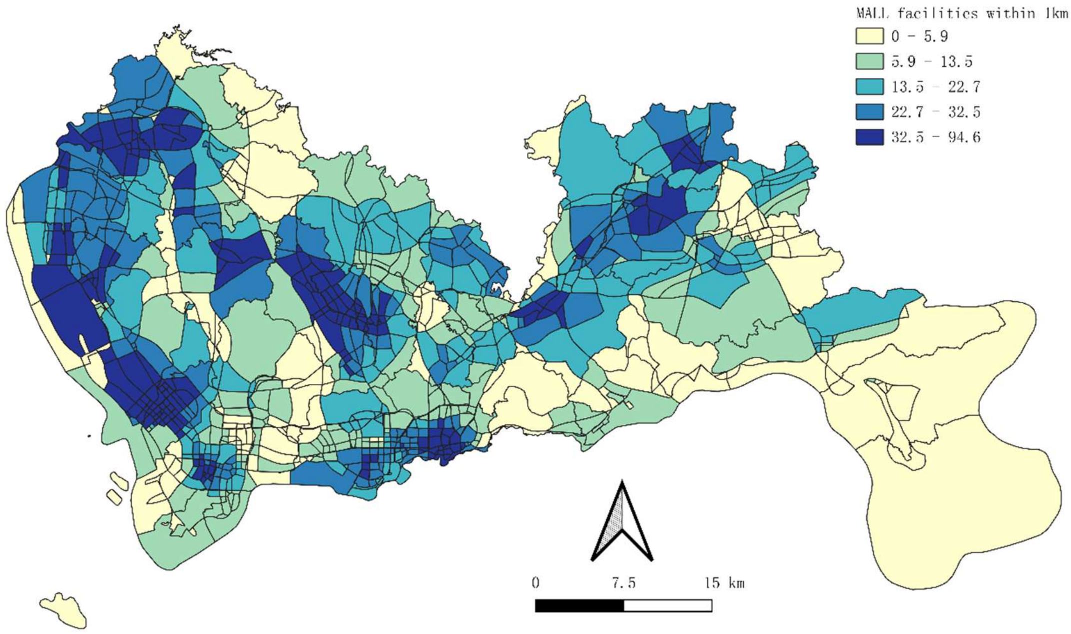

Figure 10. Average number of shopping malls within 1km radius of each neighborhood

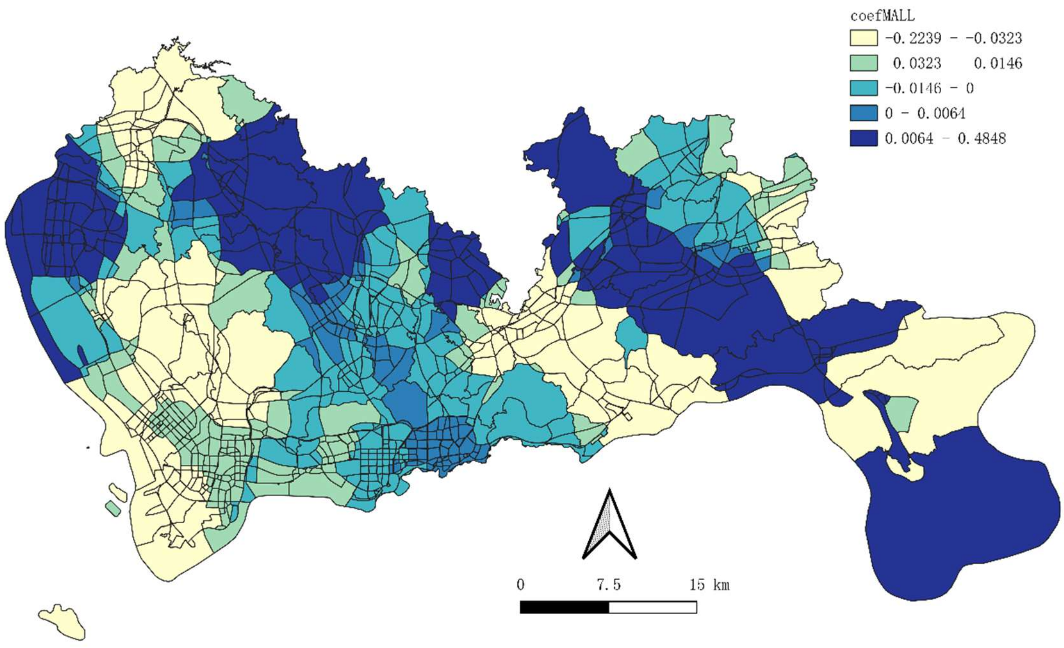

Figure 11. Distribution of coefficients (effects of shopping malls on house prices)

It can be observed from Figure 10 and 11 that shopping malls aggregate at several urban centers of different districts, while for eastern and south-east districts, accessible shopping malls for local neighborhoods are few.
The corresponding coefficients distribution show that shopping malls accessibility has greater influence on house prices in these less accessible districts. As for inner-SEZ districts, proximity to shopping malls still matters a lot and have positive influence on house price. For neighborhoods in Futian and Luohu districts, despite having been highly accessible to shopping malls, their house prices are still significantly positively influenced by such an advantage.

(3) Accessibility to skyscrapers buildings

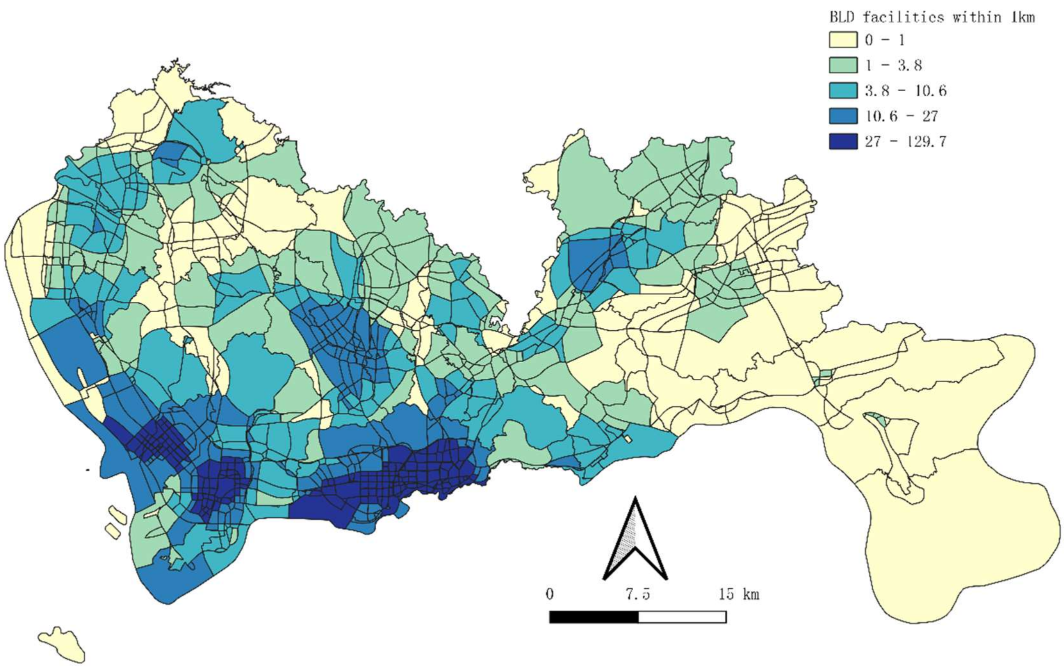

Figure 12. Average number of skyscrapers buildings within 1km radius of each neighborhood

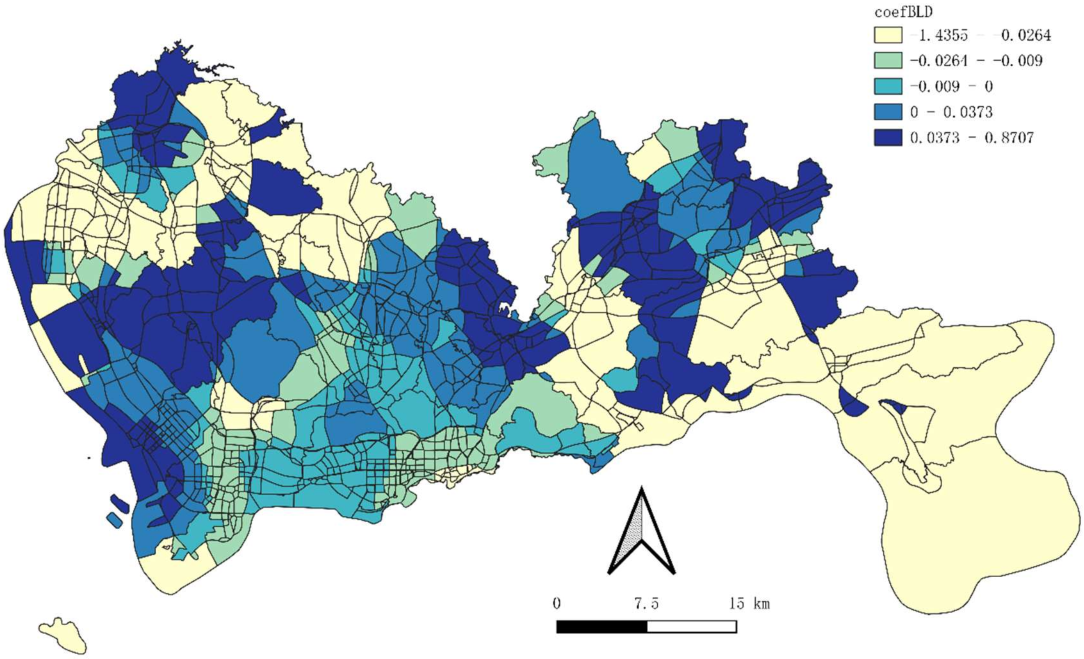

Figure 13. Distribution of coefficients (effects of skyscrapers on house prices)

It can be observed from Figure 12 and 13 that skyscrapers aggregate at inner-SEZ districts and several urban centers, while for eastern, north-east, northern and north-west districts, accessible skyscrapers for local neighborhoods are few.

The corresponding coefficients distribution show that skyscrapers accessibility has negative and weak influence on house prices in most districts. For districts that have higher numbers of aggregated skyscrapers, most of which are inner-SEZ districts, proximity to skyscrapers matters little and have negative influence on house price. For neighborhoods in outer districts, especially for those with very few proximate skyscrapers, being accessible to skyscrapers have stronger and positive influence on their house prices.

(4) Accessibility to governmental buildings

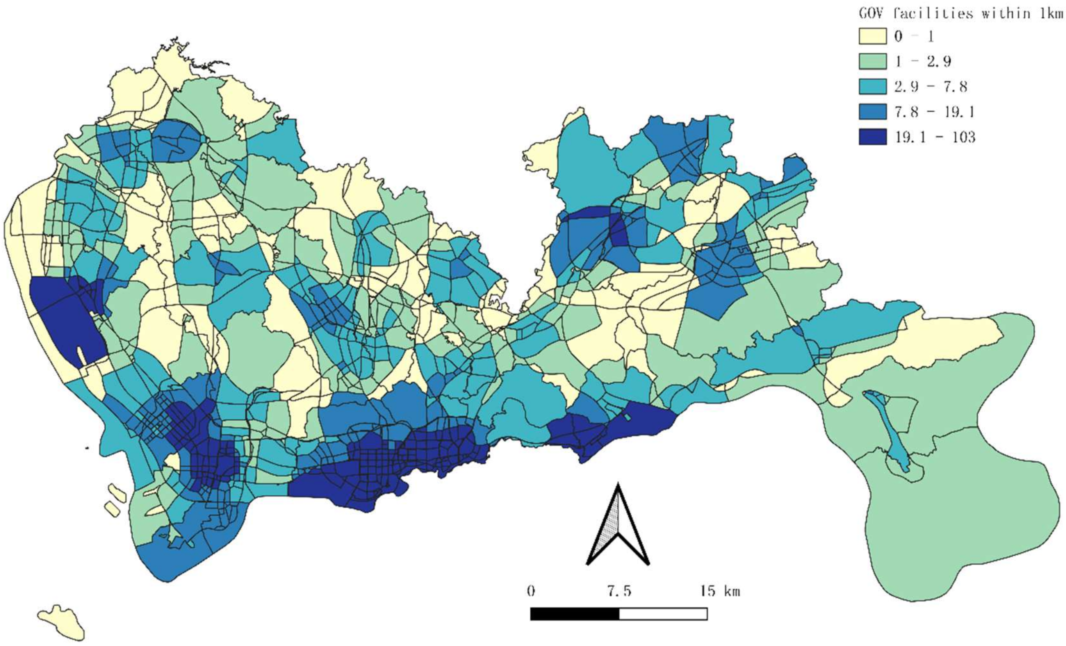

Figure 14. Average number of governmental buildings within 1km radius of each neighborhood

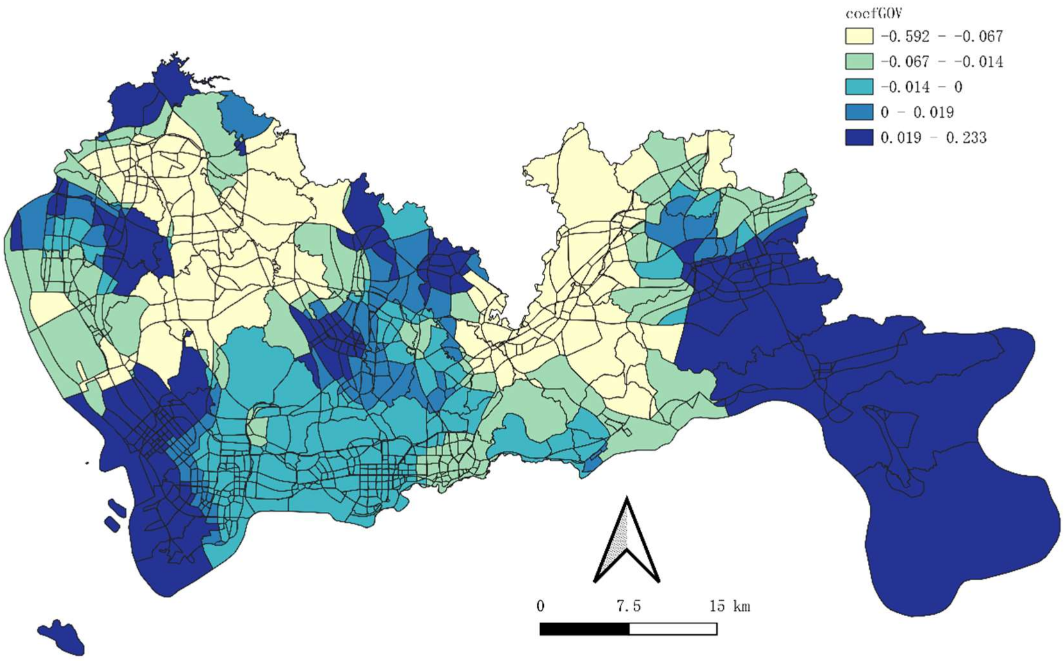

Figure 15. Distribution of coefficients (effects of governmental buildings on house prices)

It can be observed from Figure 14 and 15 that governmental buildings mainly aggregate at inner-SEZ districts and the Yantian harbor area, while for outer-SEZ districts, there is no significant aggregation of proximate governmental buildings.
The corresponding coefficients distribution show that governmental buildings accessibility has weak positive influence on house prices in inner-SEZ districts. For neighborhoods in south-west districts (Nanshan and part of Bao'an), the highly accessible governmental buildings lead to higher house prices. For neighborhoods in eastern districts (Dapeng and Pingshan), the poor accessibility to governmental buildings makes it a precious resource, resulting in higher house prices if neighborhoods are proximate to one of the governmental buildings.

(5) Accessibility to tourist spots

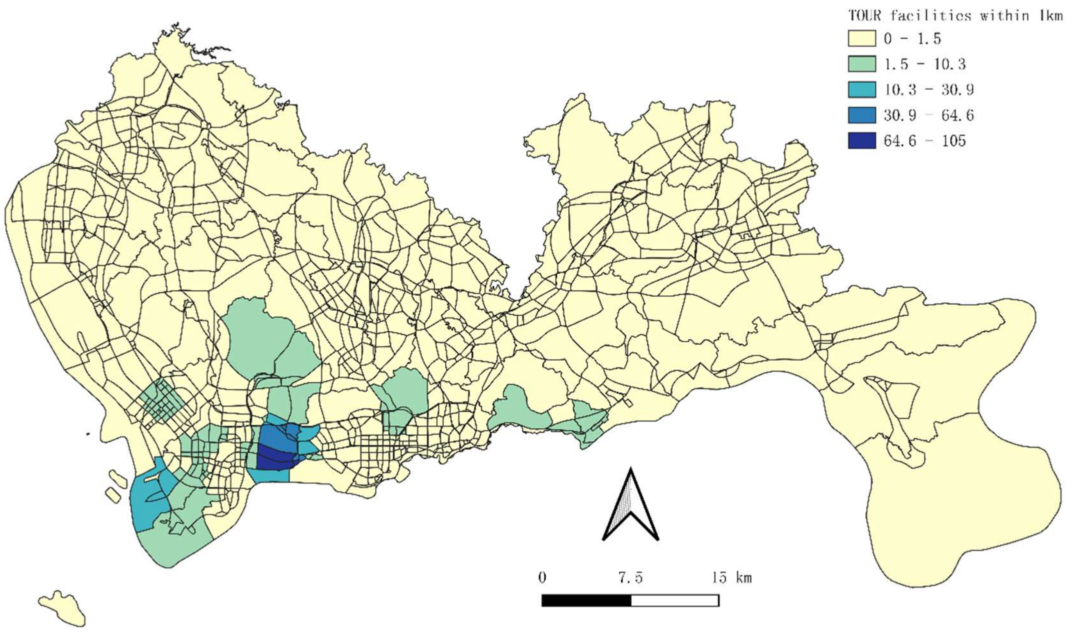

Figure 16. Average number of tourist spots within 1km radius of each neighborhood

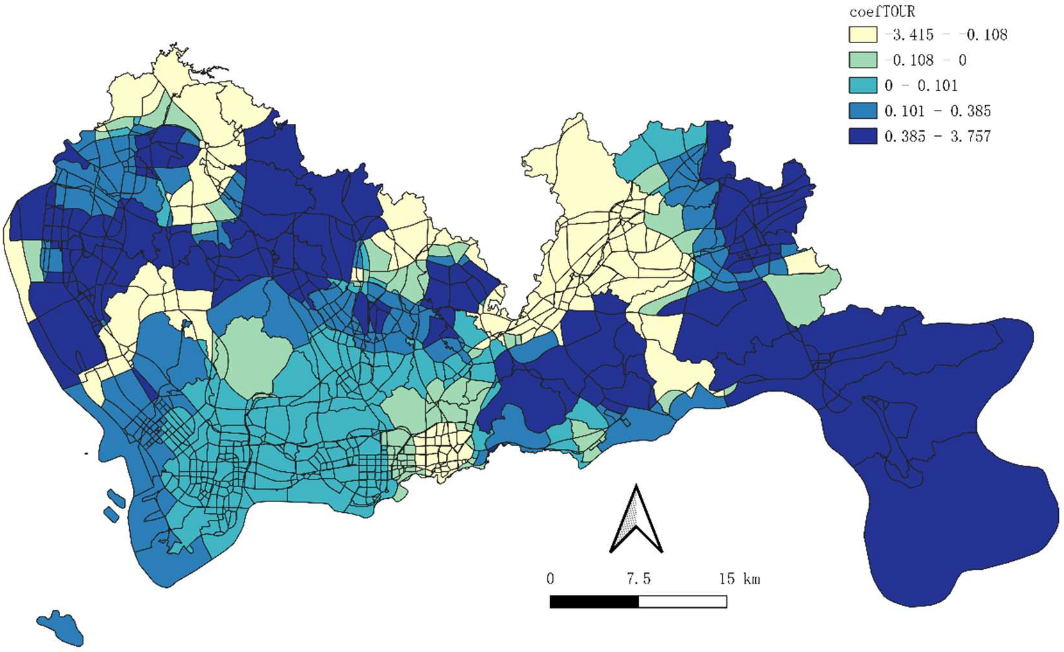

Figure 17. Distribution of coefficients (effects of tourist spots on house prices)

It can be observed from Figure 16 and 17 that tourist spots mainly aggregate at certain sub-districts (e.g. Overseas Chinese Town) and neighborhoods. For most areas, there are only 1 or 2 tourist spots or no tourist spot at all.
The corresponding coefficients distribution show that tourist spots accessibility has weak positive influence on house prices in inner-SEZ districts. For neighborhoods in outer-SEZ districts, having 1 or 2 tourist sites nearby significantly leads to higher house prices.

(6) Accessibility to metro stations

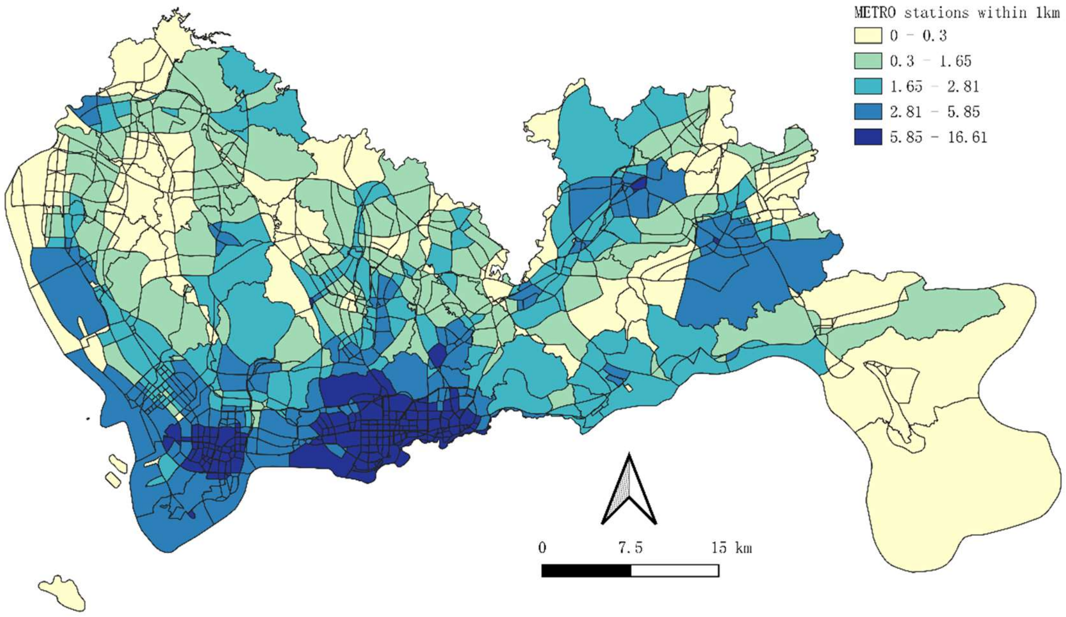

Figure 18. Average number of metro stations within 1km radius of each neighborhood

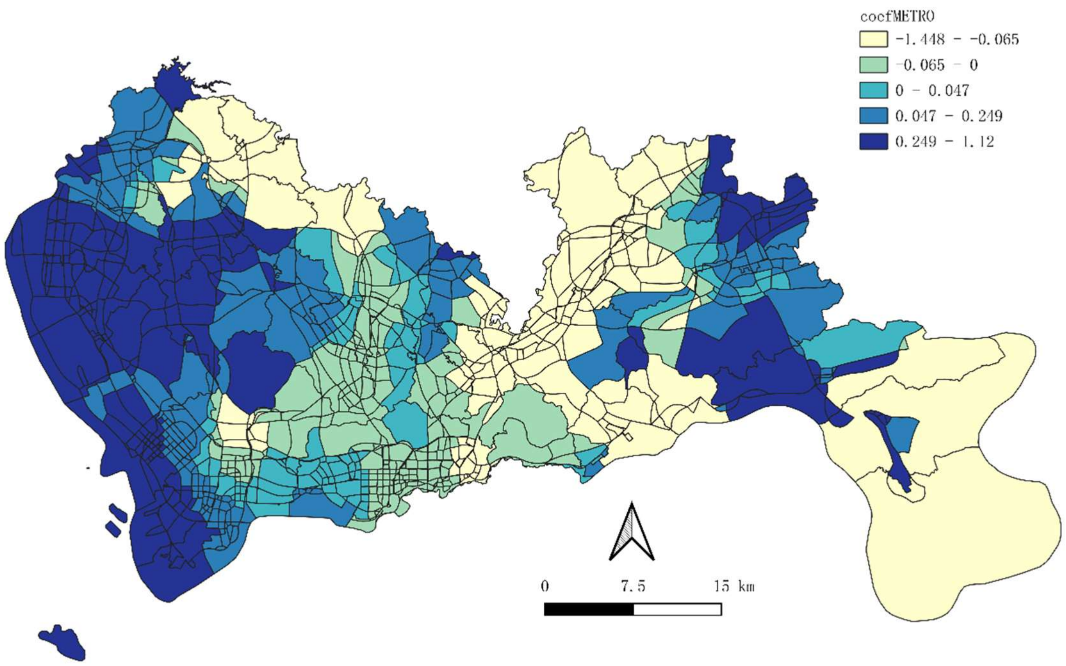

Figure 19. Distribution of coefficients (effects of metro stations on house prices)

From Figure 18 and 19 we can see that metro stations mainly aggregate at inner-SEZ districts, especially at the dense metro networks occupied by Futian central area and Luohu districts. It can be observed that accessibility to metro stations is substantially better in neighborhoods that have metro lines passing by, while for neighborhoods (most of which in outer-SEZ districts) do not proximate to metro lines, the accessibility level is significantly lower. The corresponding coefficients distribution show that metro station accessibility has positive influence on house prices in many districts, especially the western districts (Bao’an and part of Nanshan). Proximity to metro stations does not have strong influence on house prices in inner-SEZ districts such as Futian and Luohu.

(7) Accessibility to bus stations

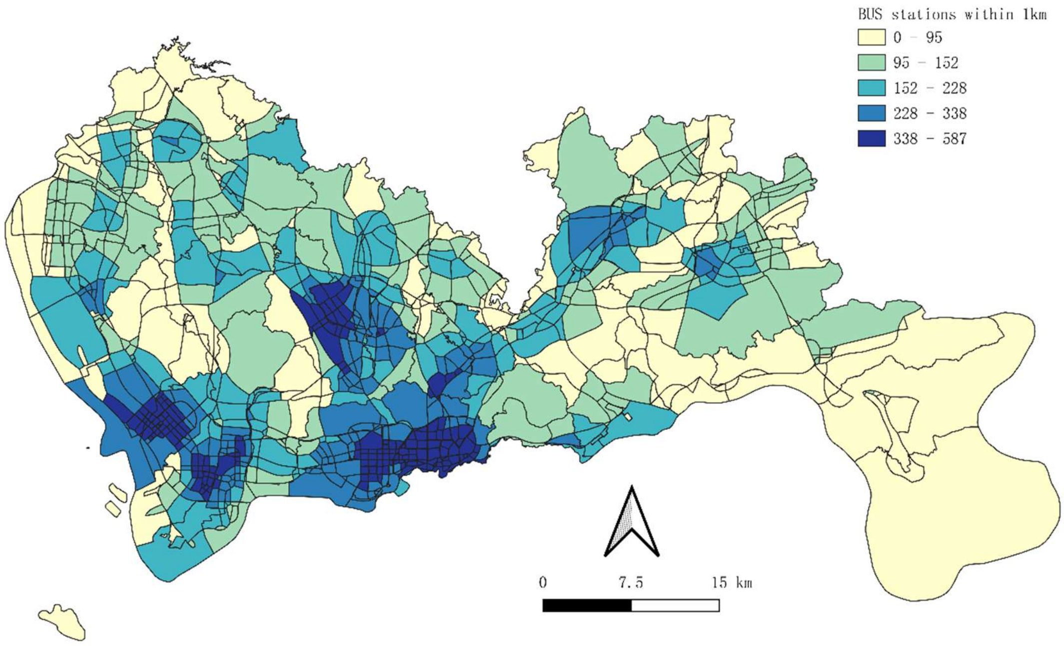

Figure 20. Average number of bus stations within 1km radius of each neighborhood

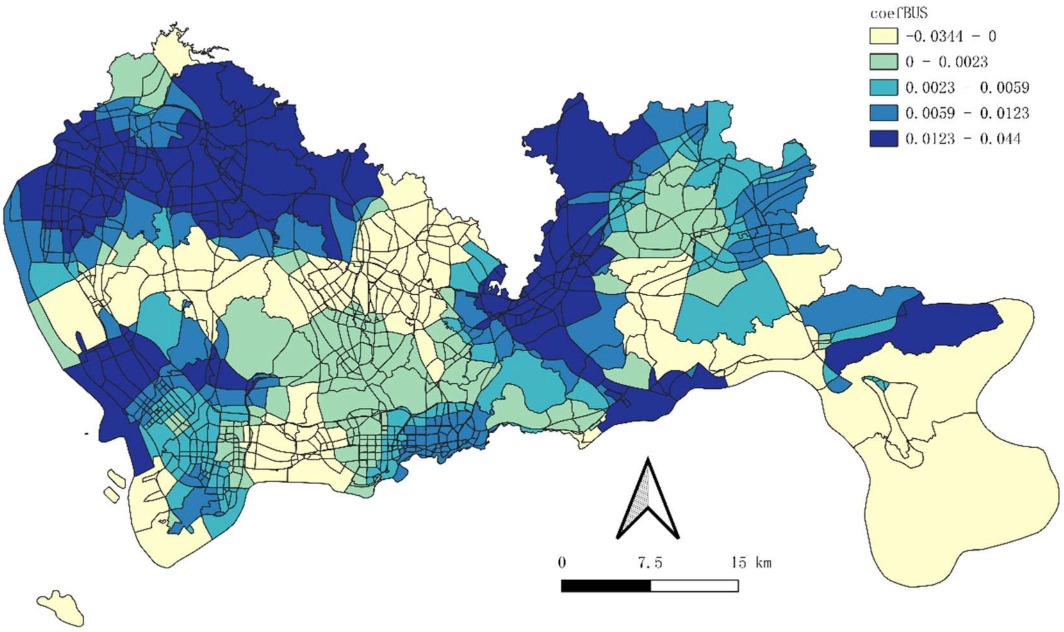

Figure 21. Distribution of coefficients (effects of bus stations on house prices)

From Figure 20 and 21 we can see that bus stations mainly aggregate at inner-SEZ districts and several sub centers. Neighborhoods with less nearby bus stations mainly situate in eastern (Dapeng) and north-west (Bao’an, Guangming) districts and areas that are not suitable for human habitat.
The corresponding coefficients distribution show that bus station accessibility has weak positive influence on house prices in many neighborhoods. The strongest influence occurs at neighborhoods in northern parts of the city where the city connects to its hinterlands. Stronger influence is also observed in the Yantian Harbor area and Bao’an airport area.

Chapter 6 Discussion

Through modelling the spatial effects of public service accessibility on neighborhoods house prices, it is found that there exist uneven distribution of accessibility to public service facilities and significant difference in their influence on house prices. The use of GWR model complement the disadvantage of the OLS model in ignoring local features and revealing the differentiate effects of each type of accessibility on different neighborhoods. The predictors which are found to have weak influence on house prices in the OLS model may actually have significant positive or negative influence on the house price regarding certain neighborhoods. The results of GWR model managed to address the spatial autocorrelation (SAC) and spatial heterogeneity (SH) within Shenzhen's distinct housing market, while the results of the OLS model fail to give rich information. This again prove the superiority of using GWR to reveal local influence of facilities in various locations (Hanink et al. 2012; Nilsson 2014).

One observation from the GWR results is that being poorly accessible to certain public service facility may result in its significancy in driving the local house price up. This phenomenon accord with the case of education facilities accessibility, shopping malls accessibility, high-rise buildings accessibility, governmental buildings accessibility, tourist spots accessibility, metro stations accessibility and bus stations accessibility. All of these facilities have positive influence on house price in most neighborhoods, especially in the neighborhoods that lack the accessibility to one of these public services consider a 1km radius (15min walk).

Another observation, is that for inner-SEZ neighborhoods, the coefficient estimates of education facilities are usually positively correlated with the predictor's value. That is, for those neighborhoods, the rich in education facilities does not soften its importance in driving up house prices. Education resource has long been a precious and costly public service for citizens in China's metropolises (Wen et al., 2018). As public primary and secondary schools are bound to certain residential area, households may wish to acquire the opportunities to enter better school districts through purchasing proximate housing, thus making education resources a strong factor in causing house price rises.

Moreover, the spatial distribution of several predictors conveys rich information consider their influence on neighborhood house prices: For instance, in most neighborhoods, there is only 1 or 2 or zero tourist spot nearby. In this case, the house prices of outer city neighborhoods show high possibility to be influenced by the proximity to tourist spots, for those sites may serve as important source of income. Another spot regarding metro stations accessibility, is that in inner-SEZ neighborhoods, being close to metro stations does not bring much influence to house prices, which is exactly the opposite situation for some outer districts. The reason for this may be that the road networks and metro lines are very densely located in the city center—take a few steps farther will bring one citizen to the next station. As for bus stations accessibility, the GWR results indicate that proximity to bus station matters for house prices of bordered neighborhoods. That result put forward a possibility that households in these neighborhoods have higher willingness to pay for intercity bus transportation, as they may have to maintain frequent connections with the hinterlands.

In general, the high level of segmentation between inner and outer SEZ, and Shenzhen' multicenter urban form enables the usage of GWR technique in this study to reflect the degree of importance of public service facilities and their accessibility. With the current results in mind, potential policy improvement could be made on finding neighborhoods that are less accessible to public service and set up targeted development plans. For households, the analysis provides a little bit of deeper understandings of what factors drive the house price in certain area, which may help them make wiser decisions in choosing residence. And most importantly, all should be aware of the spatial nonstationary of house prices and take into consideration of specific spatial characteristics affecting local housing market when making decisions.

Chapter 7 Conclusion

7.1 Major findings

Based on abundant theoretical support from the hedonic pricing theory, this study uses structural, neighborhood and locational attributes to measure the impact of public service accessibility on house price at neighborhood level. The study utilizes house price data from Anjuke and facilities POI data from Baidu Map and adopts both OLS and GWR models to investigate the influence mechanism of accessibility towards house prices at global and local levels. The major findings are follows:

The spatial distribution of public service facilities and accessibility levels in Shenzhen is uneven considering the data of 2020. The neighborhoods with high public service accessibility mainly concentrate in the inner-SEZ areas which occupied higher level of economic development, infrastructure status and residents' social status. There is an explicit gap of public service accessibility between inner-SEZ neighborhoods and outer-SEZ neighborhoods.

The key factors that drive neighborhood house prices varied over neighborhoods (spatially heterogenous). For all neighborhoods, lacking accessibility to one certain kind of public service will result in increasing importance of accessibility to that service, expressed as positive influence on the neighborhood house price. As for neighborhoods that are rich in accessibility to one certain kind of public service facilities, the importance of accessibility to that service is softened, expressed as weak or negative influence on the neighborhood house price. This phenomenon may not fit all situations, for instance, accessibility to education facilities is pursued in most neighborhoods of the city, expressed in significant positive influence on house prices.

The GWR results provide better explanatory power than the OLS model in the case of analyzing house prices in Shenzhen. As a city with explicit segmentation between inner and outer districts and multiple urban centers, the data of Shenzhen express significant spatial heterogeneity and spatial autocorrelation that require to be handled. The usage of GWR model within hedonic pricing characteristics again shows its superiority in spatial modelling of house price.

7.2 Limitations and potential improvements

Meanwhile, this study has limitations on:

The data quality for this study may be problematic as all attribute data are collected from the web. The data scraped from the Anjuke website may not be creditable enough compared to census data or governmental data regarding house price. Meanwhile, the scraping process causes some amount of data loss. The POI data may additionally encounter freshness and accuracy problems, as it is originally generated from a variety of entities and individuals without guarantee.

More explanatory variables may be required to fully catch the factors that drive house prices. This study only uses data from 3 different sources while most of them are not governmental sources. With data at smaller scales (precise to every single property) or with better statistical boundaries (spatial unit), it is expected that this study should be developed further to include a comprehensive set of structural, neighborhood and location characteristics and generate better explanatory power.

Therefore, improvements of this study will be to combine more creditable data sources, explore better ways to represent accessibility at neighborhood level, and strengthen its theoretical basis (allows for justified usage of more explanatory variables). Additionally, the house price issue of Shenzhen has far more dimensions other than public service facilities to be taken into consideration.

Reference

Adair, A. et al. (2000) ‘House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area’, *Housing Studies*, 15(5), pp. 699–716.

Alonso, W. (1964) *Location and Land Use: Toward a General Theory of Land Rent.* Harvard University Press.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Springer Netherlands.

Beckmann, M. J. (1973) ‘Equilibrium models of residential land use’, *Regional and Urban Economics*, 3(4), pp. 361–368.

Bitter, C., Mulligan, G. F. and Dall’erba, S. (2007) ‘Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method’, *Journal of Geographical Systems*, 9(1), pp. 7–27.

Bourassa, S. C., Hoesli, M. and Peng, V. S. (2003) ‘Do housing submarkets really matter?’, *Journal of Housing Economics*, 12(1), pp. 12–28.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996) ‘Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity’, *Geographical Analysis*, 28(4), pp. 281–298.

Cheshire, P. and Sheppard, S. (1997) Welfare Economics of Land Use Regulation, *Urban/Regional*. 9702001. University Library of Munich, Germany.

Choy, L. H. T., Mak, S. W. K. and Ho, W. K. O. (2007) ‘Modeling Hong Kong real estate prices’, *Journal of Housing and the Built Environment,* 22(4), pp. 359–368.

Crespo, R. and Grêt-Regamey, A. (2013) ‘Local Hedonic House-Price Modelling for Urban Planners: Advantages of Using Local Regression Techniques’, *Environment and Planning B: Planning and Design*, 40(4), pp. 664–682.

Dubin, R. A. (1992) ‘Spatial autocorrelation and neighborhood quality’, *Regional Science and Urban Economics*, 22(3), pp. 433–452.

Dubin, R. A. (1998) ‘Spatial Autocorrelation: A Primer’, *Journal of Housing Economics*, 7(4), pp. 304–327.

Evans, A. W. (1995) ‘The Property Market: Ninety Per Cent Efficient?’, *Urban Studies*, 32(1), pp. 5–29.

Fotheringham, A. S. and Brunsdon, C. (1999) ‘Local Forms of Spatial Analysis’, *Geographical Analysis*, 31(4), pp. 340–358.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2003) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Galster, G. (1996) ‘William Grigsby and the Analysis of Housing Sub-markets and Filtering’, *Urban Studies*, 33(10), pp. 1797–1805.

Gaubatz, P. (1999) ‘China’s Urban Transformation: Patterns and Processes of Morphological Change in Beijing, Shanghai and Guangzhou’, *Urban Studies*, 36(9), pp. 1495–1521.

Gibbons, S. and Machin, S. (2005) ‘Valuing rail access using transport innovations’, *Journal of Urban Economics*, 57(1), pp. 148–169.

Grigsby, W. G. (1963) *Housing Markets and Public Policy*. University of Pennsylvania Press.

Hanink, D. M., Cromley, R. G. and Ebenstein, A. Y. (2012) ‘Wage-based evidence of returns to external scale in China’s manufacturing: a spatial analysis’, *The Annals of Regional Science*, 49(1), pp. 1–16.

Helbich, M. et al. (2014) ‘Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria’, *Urban Studies*, 51(2), pp. 390–411.

Hulten, C. R. (2005) Price Hedonics: A Critical Review. *SSRN Scholarly Paper* ID 788904. Rochester, NY: Social Science Research Network.

Jim, C. Y. and Chen, W. Y. (2007) ‘Consumption preferences and environmental externalities: A hedonic analysis of the housing market in Guangzhou’, *Geoforum*, 38(2), pp. 414–431.

Lan, F. et al. (2018) ‘Spatial Effects of Public Service Facilities Accessibility on Housing Prices: A Case Study of Xi’an, China’, *Sustainability*, 10(12), p. 4503.

Lancaster, K. J. (1966) ‘A New Approach to Consumer Theory’, *Journal of Political Economy*, 74(2), pp. 132–157.

Lin, I.-H., Wu, C. and De Sousa, C. (2013) ‘Examining the economic impact of park facilities on neighboring residential property values’, *Applied Geography*, 45, pp. 322–331.

McCord, M. et al. (2012) ‘Spatial variation as a determinant of house price: Incorporating a geographically weighted regression approach within the Belfast housing market’, *Journal of Financial Management of Property and Construction*, 17(1), pp. 49–72.

McDonald, J. F. (1987) ‘The identification of urban employment subcenters’, *Journal of Urban Economics*, 21(2), pp. 242–258.

McMillen, D. P. and McDonald, J. F. (1997) ‘A nonparametric analysis of employment density in a polycentric city’, *Journal of Regional Science*, 37(4), pp. 591–612.

Melichar, J. and Kaprová, K. (2013) ‘Revealing preferences of Prague’s homebuyers toward greenery amenities: The empirical evidence of distance–size effect’, *Landscape and Urban Planning*, 109(1), pp. 56–66.

Michelson, W. M. (1977) *Environmental choice, human behavior, and residential satisfaction*. New York: Oxford University Press.

Mok, D. and Li, L.-H. (2010) ‘The Spatial Impact of Language Policies on the Marginal Bids for English Education in Hong Kong’, *Growth and Change*, 41(4), pp. 556–587.

Muth, R. F. (1969) *Cities and housing; the spatial pattern of urban residential land use*. Chicago: University of Chicago Press.

Nilsson, P. (2014) ‘Natural amenities in urban space – A geographically weighted regression approach’, *Landscape and Urban Planning*, 121, pp. 45–54.

Rosen, S. (1974) ‘Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition’, *Journal of Political Economy*, 82(1), pp. 34–55.

Straszheim, M. R. (1975) *An Econometric Analysis of the Urban Housing Market*. NBER Books. National Bureau of Economic Research, Inc.

Tobler, W. R. (1970) ‘A Computer Movie Simulating Urban Growth in the Detroit Region’, *Economic Geography*, 46, pp. 234–240.

Wen, H. et al. (2018) ‘Education quality, accessibility, and housing price: Does spatial heterogeneity exist in education capitalization?’, *Habitat International*, 78, pp. 68–82.

Wu, C. et al. (2016) ‘Spatial and Social Media Data Analytics of Housing Prices in Shenzhen, China’, *PLOS ONE*, 11(10), p. e0164553.

Xiao, Y., Orford, S. and Webster, C. J. (2016) ‘Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales’, *Environment and Planning B: Planning and Design*, 43(1), pp. 108–129.

Xu, T. (2008) ‘Heterogeneity in housing attribute prices: A study of the interaction behaviour between property specifics, location coordinates and buyers’ characteristics’, *International Journal of Housing Markets and Analysis*, 1(2), pp. 166–181.

Zhang, X. Q. (2000) ‘Privatization and the Chinese Housing Model’, *International Planning Studies*, 5(2), pp. 191–204.

Zhou, H. et al. (2011) ‘Rapid Urbanization and Implications for River Ecological Services Restoration: Case Study in Shenzhen, China’, *Journal of Urban Planning and Development*, 137(2), pp. 121–132.