IHS is the international institute of urban management of Erasmus University Rotterdam

IHS
Making cities work

MSc Programme in Urban Management and Development

Rotterdam, The Netherlands

September 2018

Thesis

Title: Incorporating spatial heterogeneity in determinants of second-hand housing prices in Zhengzhou City

Name: Chunyue Mao

Supervisor: Drs. J. (Jeroen) van Haaren

Organization: Erasmus Centre for Urban Port and Transport Economics, Erasmus University Rotterdam

Specialization: Urban Competitiveness and Resilience

UMD 14

MASTER'S PROGRAMME IN URBAN MANAGEMENT AND DEVELOPMENT

(October 2017 – September 2018)

Incorporating spatial heterogeneity in determinants of second-hand housing prices in Zhengzhou City

Chunyue Mao

China

Supervisor: Drs. J. (Jeroen) van Haaren

UMD 14 Report number: 1140

Rotterdam, September 2018

Summary

This paper demonstrates how to construct an effective and accurate real estate appraisal method using a geographically weighted hedonic regression equation. It considers the spatial distribution of second-hand residential houses and reveals the mechanism of various determinants impacting on houses by explicitly incorporating spatial heterogeneity. The paper shows that space matters and implicit valuation of housing attributes differs over space. Also, the results suggest that incorporating space explicitly in real estate valuation models will improve the accuracy of property appraisal.

With the rapid development of China's real estate industry, traditional single asset valuation methods cannot meet the demand appraisal of properties. Combined with the impending introduction of property taxes, mass appraisal of real estate is needed. However, current property valuation techniques mainly rely on the 'Land Datum Value Method' and the 'Market Comparison Approach', which are both prone to inaccuracy and subjective opinions of the appraisers. Actually, in active real estate markets, an estimation of the relationship between housing determinants and the housing price requires rich samples of housing price. Therefore, a first step would be to quantitatively describe this relationship by means of geostatistical models, and provide an objective appraisal method for housing price. Besides, another problem imbedded in the current valuation methods is the lack of consideration for spatial heterogeneity. Residential properties are highly heterogeneous good, by virtue of their differing location. Even at the same time, prices vary greatly depending on the different structural, neighborhood and locational characteristics over space. However, existing real estate appraisal methods do not take spatial heterogeneity into account, resulting inaccurate assessment.

This paper firstly analyses the research background including the national real estate market and the Zhengzhou housing market. Subsequently, previous studies on the Hedonic Price Model and spatial heterogeneity of housing prices are discussed. Next, the core empirical part of this research is presented. In this paper, housing attributes and prices for 8864 residential houses in Zhengzhou City analysed. Firstly, some spatial features are derived based on the Exploratory Spatial Data Analysis (ESDA) method in ArcGIS. Secondly, an Ordinary Least-squares (OLS) estimation based on the Hedonic Price Model is presented to explain the basic correlation between the housing determinants and housing prices. Thirdly, the Global Moran's Index is evaluated to examine if there is spatial autocorrelation problem in the OLS estimation. After that, a Geographically Weighted Regression (GWR) estimation is used to test the spatial heterogeneity in the price composition of the second-hand housing market in Zhengzhou City. The final step is to compare these two estimation methods and the conclusion can be drawn as follows:

Compared with the OLS estimation, the GWR estimation has a better performance in the adjusted R^2, AICc value, the Moran's Index and sum squares of the residuals, indicating that the GWR estimation is more suitable for the evaluation of housing price. In other words, space matters in determining house prices and attributes of houses are valued differently over space. Also, the coefficients in the GWR estimation of these housing determinants vary in space with their different spatial locations and reflects the spatial heterogeneity of the influence of these indicators. This would indicate that cities follow particular patterns with respect to their morphology, which drives the difference in prices of housing attributes. These results suggest that, taking the spatial heterogeneity of housing price into account, will provide more effective and accurate models for the housing price evaluation. Particularly with respect to mass appraisal methods, incorporating spatial heterogeneity will result in more accurate estimates. Taxation departments can use the results of this paper to improve their computer assisted valuation models. Nevertheless, additional research may be needed, for instance to look into the questions of poly-centricity, as well as addressing current data (and computation) limitations.

Keywords

Real estate appraisal, Urban housing price, Housing determinants, Spatial heterogeneity, Geographically Weighted Regression (GWR)

Acknowledgements

It is exciting to say that after six months of effort, this master paper has been finished. I can hardly believe that time flies so fast and the whole year of my study in the Netherlands is almost finished.

First and foremost, I want to show my gratitude to IHS. I can still recall my memory of the happiness when I received the e-mail of the admission letter from IHS. This is my first time to study abroad and I am so happy to meet so many friends, teachers and experience a brand-new culture.

I would like to convey my gratefulness to my supervisor Jeroen van Haaren for his patient and constant supervision. He shared sparkling ideas with me checked every detail of my research. There is an old saying in China and here I want to translate that to show my thankfulness to my supervisor. “Give a man a fish and you feed him for a day. But teach him how to fish and you feed him for a lifetime”. Actually, I honour him so much is not only because the detailed comments and recommendations he gave me, but the thought to think more logical and to seek the practical value of a thesis. These minds will support me to go deeper into my future researches and “feed” me for a lifetime.

I also want to thank my supervisor in China, Professor Weidong Qu. I cannot finish this thesis without the data of Zhengzhou City that he provided for me. I also discussed with him a lot about my study life in the Netherlands and he gave me continuing encouragements and academic comments. Also, I own my gratefulness to Mr. Bardia Mashhoodi for his kindly guidance on the GWR estimation. Thanks for his excellent lecture about the GWR estimation and also the continuous support to let me go deeper and deal with some problems in my thesis.

I would like to thank all the staff in UCR and the IHS. Thanks for Paula and Ronald for their hard work and technical guidance throughout my study. I also want to express my gratitude to two research assistants, Rupinder and Indriany, who also gave me useful feedback and comments to improve my thesis. Same thanks go for other staff in HIS for their support and help.

Besides, I own my gratitude to my parents. They gave me the strongest support, encouraged and trusted me all the time. They provide me this opportunity to open my mind and taught me to have strong belief in myself. Thank you for always standing by my side.

Finally, I would like to express my gratefulness to my friends and colleagues, especially Mingming, Jing, Renjia, Bingying, Chaoyuan, Yujia, Qi, Jianhua, Bo and Jimmy. Your involvement and companion have contributed a lot for the finishing of my thesis. I really miss the time we spent together, cooking, travelling and learning. Also, I am grateful to my foreign friends, you all gave me a lot of help and confidence in the whole year.

Finally, thanks for all the people I met and the memory they left during my master study in the UCR.

Abbreviations

AIC	Akaike Information Criterion
CBD	Central Business District
ESDA	Exploratory Spatial Data Analysis
GIS	Geographic Information System
GWR	Geographically Weighted Regression
IDW	Inverse Distance Weighted
IHS	Institute for Housing and Urban Development
MWR	Moving Window Regression
OLS	Ordinary Least-squares
SEM	Spatial Expansion Method
VIF	Variance Inflation Factor

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background

1.1.1 Housing system reform in China

Prior to the 1980s, all the houses were established by Communist work units or by residents in China. However, China transformed its institutions from state-planned housing arrangement to a preliminary market-oriented housing system in 1983 (Y. P. Wang and Murie, 2000), which has had important consequences for the Chinese housing market. Moreover, the commercialization of land and real estate marked the conversion to a modern housing system. The main objective of this reform is to decentralize power to local governments. In addition, land value is now determined in urban areas through trading. Since 1993, the Chinese government began to implement institutions and policies on the real estate industry so that the housing market could be controlled and curbed to some extent. In 1998, the central government eliminated the welfare housing distribution system and established market-oriented mechanism (Y. P. Wang and Murie, 2000). During the period of 1998 to 2003, a multi-tier urban housing supply system was built which encompassed commercial housing construction, sales, purchases, and leasing activities. The real estate market was still at a developing stage during this period, without an efficiently functioning market.

From the year of 2003, investment in the housing market overheated in many provinces in China. The housing prices in Shanghai, Beijing, Nanjing showed a strong upward trend, ranking on the top of the list. Since then, housing prices in other major cities across the country

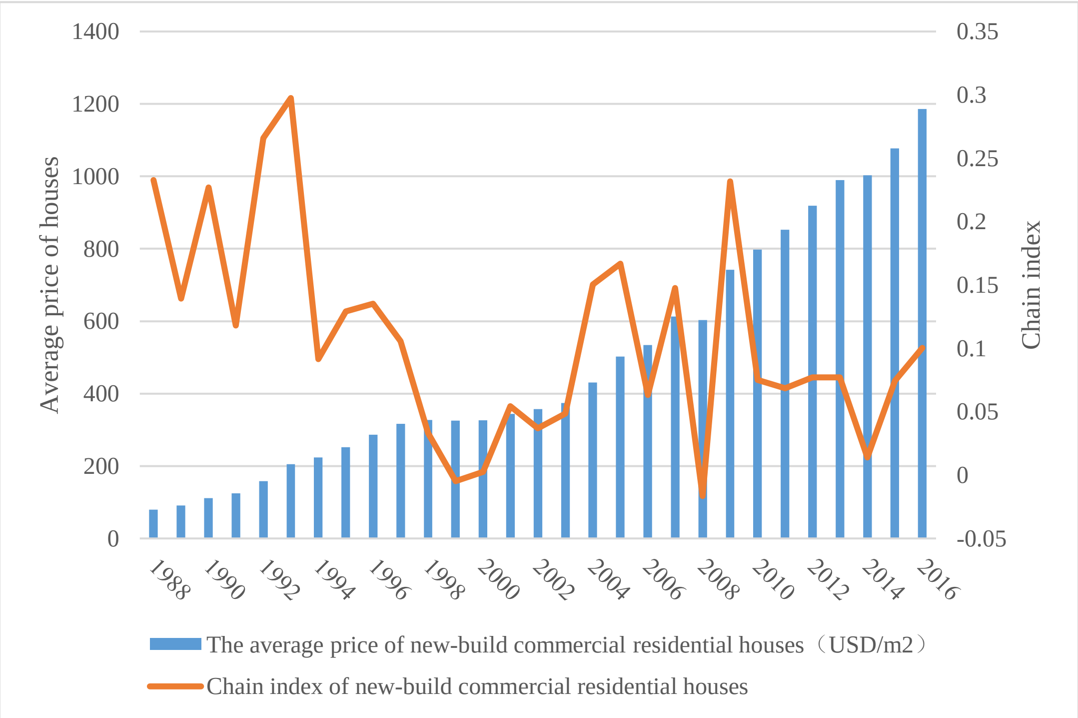

Figure 1 The trend of housing price in China from 1988 to 2016 (Source: the author based on data from the China Statistical Yearbooks from 1989 to 2017)

also increased rapidly. Figure 1, shows both the average price and chain index[1] of commercial residential houses from 1988 to 2016. It provides a clear overview of the snowballing

[1] The chain index = ($price_{t1}$-$price_{t0}$)/ $price_{t0}$. If the chain index is more than 0, it means the housing price has a upward trend.

development of the housing market in China, with strong year on year growth. Also, Figure 1 shows a sharp slowdown in the increasing of new-build houses can be seen from 2010 coupled with the same tendency of Chinese economic growth.

However, the primary commercial residential housing market has accumulated a considerable (existing) housing stock after more than two decades of development. Combined with sustained rural-urban migration flows in the labor force, a thriving housing market for existing homes has emerged (Zhu, Tang, et al., 2014). Statistics show that the aggregate transaction price of existing homes in the first-tier cities has surpassed 475 billion USD[2] in 2016, double that of newly-build homes. This is a clear indication that China's real estate industry has shifted from the emerging real estate market to a mature market with a considerable housing stock. Newly build housing provides additions to this stock, as well as replaces depreciated homes.

With the reform and improvement of the housing system, the housing market in China will become more active and real estate appraisal activities will act an increasingly imperative role in the real estate market. Because each of the real estate transfer, leasing, mortgage, taxation and the compensation for expropriation need to be based on the assessment the valued of real estate. This research paper therefore aims to contribute to a place specific real estate appraisal model for the housing market in China.

1.1.2 State of the secondary housing market in Zhengzhou City

Zhengzhou City is the capital city of Henan Province in east-central China, with a population of around 9.57 million (National Bureau of Statistics of China, 2016). The total area of Zhengzhou City is 7446 square kilometers, and there are 8 jurisdictions and one direct management county in the city. From a historical perspective, it is a typical representative of Chinese Great Ancient Capitals. Zhengzhou City is a rapidly growing city. As the political, economic, cultural and financial center of Henan Province, Zhengzhou City is an important regional center and significant transportation hubs in the central region of China.

During the first phase of housing reform in 1992, the incorporation of a local private real estate developer in Henan Province marked the first step of real estate development in Zhengzhou. At the beginning of the 21st century, the concept of 'the rise of Central China' promoted the development of Zhengdong New District which attracted more and more people to Zhengzhou City. As a result, the scene of this economic boom resulted in a strong demand for houses for people who worked in Zhengdong New district. With the accelerated rate of urbanization, the construction of Zhengdong New District and the transformation of the old industrial district in western Zhengzhou, investors are motivated by high profits from the residential housing market that also underpinned the continuous rise in housing prices. Since prices of new-build houses in Zhengzhou City are very high, the second-hand houses are increasingly preferred by more and more prospective home owners.

In 2017, the average selling price of existing homes was USD 1681 per m^2 with a total transaction area of just over six million square meter(6.135) (see Figure 2). Faced with an ever-increasing trading volume and the upward trend of prices in residential housing markets, research on residential houses price have been one of the main topics of academic debate. However, most of the existing studies only focused on the influence of various factors without exploring the spatial heterogeneity in determinants of urban housing prices (Wen, 2010, Tang, 2012).

1.2 Problem Statement

[2] Exchange rate of USD against RMB (CNY): 1:6.3 (30.03.2018).

The residential housing market has grown substantially over the last twenty years and it has become one of the main pillars of the economy of China (Zhong, 2012). As the stock of housing has steadily increased over the years, the market for existing homes developed into substantial

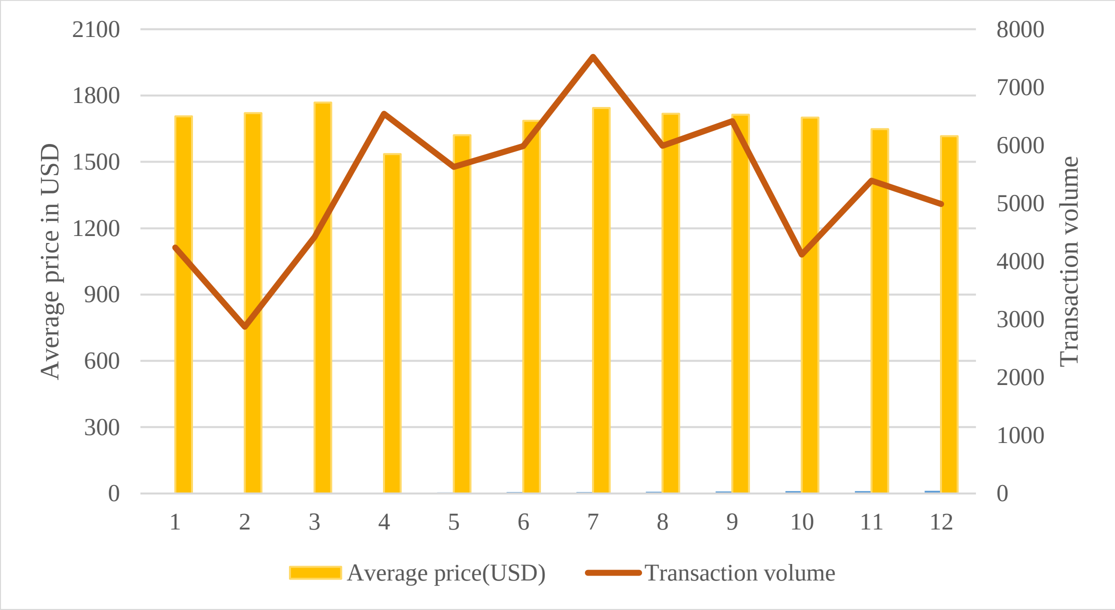

Figure 2 The monthly average price and transaction volume of second-hand housing in Zhengzhou City in 2017 (Source: the author based on data from the Housing security and Housing Authority of Zhengzhou City)

proportions. As witnessed in other developed countries, the process of urbanization is conducive to maturation of second-hand housing markets in cities. Even though China's second-hand real estate market has begun to take shape, a considerable gap between the domestic second-hand housing market and the mature real estate market in other developed countries exists (Rong, 2014). Nevertheless, China is rapidly closing this gap. Given an average life-span of 70 years for residential buildings, we may expect China’s housing market to be fully focused on replacement within the next decades. Given the increasing importance of existing houses over newly build dwellings, additional research into second-hand markets is highly interesting.

Most of the available contributions on Chinese markets for existing homes cover the real estate market of second-hand housing in mega cities, such as Beijing, Shanghai or Shenzhen, whereas only few are provided for second-tier cities. Also, compared with the fluctuated housing markets with many housing bubbles and idiosyncratic characteristics in the first-tier cities, the housing market in many second-tier cities has remained steady from both the supply side and demand side till now. Therefore, this paper is contributed to fill a gap in existing literatures in Chinese housing market by focusing on second-tier cities. In this paper Zhengzhou City is selected as a representative second-tier city in China.

From the micro-economic point of view, the analysis of urban second-hand housing prices is based on Hedonic Price Model, in which house prices are broken down into single factors (Z. Li and Wu, 2008). To be specific, traditional empirical analysis focusses on the structural, locational and neighbourhood characteristics to measure the implicit prices of attributes of existing homes. These characteristics, combined with constant quality price indices make up transaction prices in markets for existing homes. In its most elementary form the Hedonic Price Model assumes a constant price structure of all real estate attributes within the entire research area, which may cause biases of the regression coefficients if price formation processes are heterogeneous between areas (Bitter, Mulligan, et al., 2007). It is sensible to assume that price formation processes differ between areas within the city. Therefore, this paper takes the spatial variation in determinants of second-hand housing prices into account. By using Geographically Weighted Regression (GWR) estimation technique, this paper will analyze the spatial distribution of second-hand residential houses in Zhengzhou City and reveal the mechanism of various determinants impacting on houses.

The current real estate tax system in China has the reality of "heavy tax in the circulation part instead of owing part". There are only a few types of retained taxes and the sales tax revenue accounts for more than 90% of the total real estate tax revenue, which shows an upward trend year by year(Zou, Wu, et al., 2009). Moreover, as the tax base of each real estate sales tax is the actual transaction price, actors may be incentivised to set it below market prices, seeking alternative forms of compensation (for instance by simultaneously trading consumer goods at inflated prices). Therefore, transaction prices in the real estate market may in some cases not reflect actual market prices for housing. It is challenging to assess the deviation between the transaction prices and the market prices. This paper aims to provide a scientific and reasonable evaluation of the real estate market value, as a key to the levy of various real estate taxes. Also, since the method of real estate appraisal in China is mainly focus on the Market Comparison Approach which is based on the large amount of transaction cases of stock housing and contains subjective opinions of the appraisers. In this light, the conclusion of this research will provide input for a more effective and accurate model for taxation departments.

1.3 Research Objectives

Based on the analysis above, the objectives of this study are set up as follows:

a. To explore what microeconomic factors and how these determinants affect the price of second-hand houses in Zhengzhou City.
b. To examine the spatial heterogeneity in determinants of second-hand housing price in Zhengzhou City.

1.4 Research Question

The main question is:

What are the determinants of second-hand housing prices and do these determinants vary between different areas?

The sub-questions are:

a. Based on Hedonic Price Model, what factors and how they will influence the second-hand housing prices in Zhengzhou City?
b. Is there spatial heterogeneity in the price composition of the second-hand housing market in Zhengzhou City?

1.5 Significance of the Study

1.5.1 Academic significance

(1) Provides basis for real estate price analysis

Through the establishment of the Hedonic Price Model, it is possible to better analyze those key factors that influence second-hand housing prices and provide a theoretical basis for real estate price analysis. Second-hand housing prices are related to the macro-economy and the people's livelihood. As a result, operating scientific and effective analytical tools has a crucial role in forecasting housing prices and lays a solid foundation for establishing a real estate price analysis system.

(2) Spatial econometrical analysis

The Hedonic Price Model is a well-established analytical framework for studying the microscopic characteristics of houses (McCord, Davis, et al., 2012). However, it lacks consideration of spatial effects. Spatial econometrics breaks the assumption that the influence of each determinant is stationary within the entire target area, without taking the spatial variation into account. On the basis of the GIS, the latitude and longitude coordinates of each dwelling are collected, data including latitude and longitude coordinates are established through GIS, and spatial econometric methods are introduced to analyze the systematic variation in housing prices and reduce the heteroskedastic error terms in the model. The neighborhood effect can also be eliminated and more reasonable results can be obtained.

1.5.2 Practical significance

In terms of practical relevance, based on the Hedonic Price Model, this paper analyzes the characteristic variables that affect urban second-hand housing prices and quantifies the influence of these factors, which can provide consumers with certain guidance for the decision-making of housing purchase. At the same time, considering the existence of the spatial heterogeneity of housing prices, this paper offers insight into how these factors in Hedonic Price Model are determined by the spatial effect. Therefore, it can provide insight into improving upon the accuracy the model of real estate assessment. Simultaneously, as the existing method of real estate appraisal in China is based on the stock housing, if there is no sufficient transaction case in one area, it is difficult to apply this method. However, the Hedonic Price Model is targeted to build the functional relationship between the characteristics of the property and its value. Regardless of whether or not transaction cases exist in the region, this relationship is constant between attributes, but possibly varies over space. This paper provides a methodology to assess these relationships, which greatly broadens the scope and feasibility of real estate assessment. If the model can be established in combination with the actual conditions in China, it will be of great significance to the levy of real estate tax in China.

1.6 Scope and Limitation

Basically, combining the spatial analysis of GWR and the classical Hedonic Price Model, this paper will carry out a systematic study of the determinants of second-hand housing price in Zhengzhou City and also take the spatial heterogeneity into account. By applying the means of spatial economics, it will provide a method to visualize the spatial variation of second-hand houses in Zhengzhou City and an insight into the mechanism of the price-making process.

Also, the limitations of this study could be analysed as follows:

(1) Beyond the scope of this study, other determinants can be rendered for this framework, such as the building equality and behavioural habits of inhabitants within each neighborhood. Nevertheless, to distinguish if these factors are omitted in models, which could cause omitted variable bias, more computation-efficient methods are planned to be used to test the models. Moreover, innumerable researchers have illustrated that omitted variables affect spatial regression methods less severely than ordinary least-squares (OLS) estimations (Dubin, 1988, Brasington and Hite, 2005, Pace and LeSage, 2010).

(2) This research is mainly focused on the price of second-hand housing in five urban districts in Zhengzhou City. With the development of Zhengdong New District, the average price of second-hand houses there is already opened the gap from the other districts in Zhengzhou City. Also, the impact of Zhengdong New District will probably decrease the accuracy of the analysis.

(3) Because real estate sales cannot be measured at the same time point, the lack of analysis on

time series within a year is still a limitation of this paper. At the same time, since the limitation of the database, this study cannot be carried out with a continuous time-bound, which means the impact of time should be the next factor to be considered. Also, the endogeneity problem caused by cross-sectional data which refers to the circulated impact of the spillover effect of housing price on spatial heterogeneous coefficients of different indicators needs to be considered in the future.

Chapter 2: Literature Review / Theory

Housing price is very much related to the healthy macroeconomic development and the happiness of residents within a country. Thus, numerous researches have analysed the relationship between housing determinants and the residential housing prices based on the theoretical Hedonic Price Model and different statistical models.

This literature review chapter will be organized into six parts. The first part will briefly introduce second-hand housing. The second part will display the historical development of the Hedonic Price Model. After that, the third part will go into the applications of the Hedonic Price Model based on the previous related research. The fourth part focus on analyzing the spatial heterogeneity in attributes of housing prices. Also, there is a concise summary of the whole literature review in the fifth part. Finally, the conceptual framework will be built in the last part.

2.1 Second-hand Housing

2.1.1 Definition of second-hand housing

The second-hand housing, compared with the new-build housing, refers to the property that has already been registered and owned by an individual and the ownership is going to be re-transferred to others (Shen, 2015). In general, properties with clear housing property right certificates and transactions after the first sale are called second-hand housing.

2.1.2 Advantages of second-hand housing

The advantages of the second-hand housing can be divided into three aspects. Firstly, the auxiliary facilities located around the second-hand housing are well-designed, integrating living, business, and service demand into account. Secondly, the quality of second-hand housing has been tested compared with a new-build housing. After several years of use, if the house has problems, consumers can easily identify that. Besides, there are less pollution sources in the furnished second-hand houses which avoid the decoration pollution in new houses. The final advantage of second-hand houses is the relatively cheap price, which are better choices for more middle-low income salary earners (Wu, 2016). Based on all these advantages, second-hand houses have already been a cheap and comfortable alternative for many peoples.

2.2 The Theoretical Development of Hedonic Price Model

2.2.1 The theoretical framework of Hedonic Price Model

Etymologically speaking, the word "hedonic" stems from the Greek word hedonikos, which refers to pleasure. Within the economic background, “hedonic” means the utility or satisfaction after the consumption of one specific good or service (K. W. Chau and Chin, 2002).

Generally, Hedonic Price Model was developed and introduced into pricing mechanism long before the conceptual and analytical lexicon was established (Triplett, 1986). Studies on Hedonic Price Model can be traced back to the period between the 1920s and 1930s. According to Waugh (1928), he applied Hedonic Price equation to analyze the function between vegetable quality differences and prices in Boston, which is the earliest research on exploring the influence of commodity characteristics exerting on housing price. Court (1939) also used this model to explore the impact of car attributes on size, weight, fuel consumption and so on. However, a theoretical framework was not built according to the studies in this period.

After a period of silence, the price has achieved significantly theoretical development in the 1960s and 1970s. Since Griliches’s (1961) research of the compilation of automobile price index and another important study of Rosen (1974) on emphasizing the relationship between attributes of goods and willingness of payment, Hedonic Price Model has come into vogue in the economics. After that, scholars continued to conduct large numbers of research and introduced the analysis and related technologies to deal with heterogeneous commodity issues. However, it is generally recognized among academia that the Hedonic Price Model was rooted in Lancaster's (1966) consumer theory and Rosen's (1974) theoretical model (K. W. Chau and Chin, 2002). According to the new consumer theory of Lancaster (1966), the price of a commodity is controlled by the characteristics of itself rather than the good, which provides the theoretical basis for the Microeconomic Theory. Based on the study of Lancaster, Rosen proposed a specific Hedonic Price Model for residential properties and it was widely used in the study of the relationship between housing prices and living environment (Chen, 2009). In a nutshell, all of these researches were aimed at building the connection between the observed prices of miscellaneous commodities and their myriad of characteristics.

Since the 1980s, econometrics has greatly promoted the development of Hedonic Price Model, making it gradually mature in theory and technology. From then on, empirical researches based on Hedonic Price Model in the field of real estate have become widespread. Most of the empirical studies in the 1980s focused on the establishment of housing price indexes. Since the 1990s, especially the beginning of the 21st century, with the continuous emergence of concepts such as ecological housing, healthy housing, and green housing, also the prevalence of humanitarianism, scholars imputed more factors into the housing price analysis. To be specific, researchers considered the impact of the housing attributes from the perspectives of infrastructure and ecological environment. Furthermore, traffic conditions, educational facilities and public and green spaces were included in the infrastructure conditions of a house (Chen, 2009).

2.2.2 The Hedonic Price Model empirical issues

In practice, the choice of the functional form of the Hedonic Price Model has been served as a major empirical issue at hand. Basically, there are three kinds of popular functional forms, such as linear functional form, semi-logarithmic functional form and log-log functional form which could be frequently seen in a Hedonic Price Model (Sirmans, Macpherson, et al., 2005). Researchers usually compare different functional forms and use the Likelihood ratio tests to examine which appropriate functional form should be applied in the study since the failure to use an accurate choice may result in inconstant estimates (Jensen, Liljemark, et al., 1981, Halvorsen and Pollakowski, 1981). Besides, the Likelihood ratio, stemmed from the Box-Cox transformation can improve the fitness of the data (Rasmussen and Zuehlke, 1990). Box-Cox transformation is a kind of data transformation, frequently used in statistical modelling when continuous response variables do not distribute normally. As a result, Box-Cox transformation is aimed at reducing the correlation between unobservable error and predictor to some extent.

However, Box-Cox is criticized to reduce the accuracy of each of the coefficients in the model and the fail to lead to more precise estimates (Cassel and Mendelsohn, 1985). What is more, a further shortcoming of Box-Cox transformation was indicated that it did not allow binary or dummy variables in the model as their impact are not firmly positive (Linneman, 1980, So, Tse, et al., 1997). In general, although the functional form is very significant to the Hedonic Price Model, there is still a lack of guidance on selecting the appropriate equation form in the long research history (Halvorsen and Pollakowski, 1981, Butler, 1982).

Some scholars also introduced algorithms in other fields into the Hedonic Price Model. For instance, Peterson and Flanagan (2009) integrated Artificial Neural Network Analysis and GIS spatial analysis techniques into the Hedonic Price Model when studying real estate value factors, in order to improve the evaluation effect. The conclusion showed that ANN allowed including dummy variables which provided a better empirical form for the conceptual Hedonic Price Model. LeSage and Pace (2004) and Bitter (2007) introduced spatial autocorrelation and spatial error analysis into the Hedonic Price Model for estimating the spatial heterogeneity of the residential housing price.

2.2.3 The Hedonic Price Model empirical assumptions

In light of the Hedonic Price Model theory, consumers' demand for heterogeneous commodities is not determined by the product itself, but on the characteristics or attributes which are inherent in the product. Consumers transform the “input” through buying and using commodities into utilities and the level of utility depends on the bundle of groceries with commodities. For the convenience of the analysis, when the model is established, it is generally assumed that certain assumptions are made (Chen, 2009). Two core assumptions are contributed significantly to the empirical framework of the Hedonic Price Model as follows:

The first assumption is the heterogeneity of commodities. The theoretical basis of the Hedonic Price Model is to look into the heterogeneity of products and the conclusions are also provided for goods with great variation, such as housing products, automobiles and labor. The heterogeneity of commodities provides each commodity with characteristics that are different from other similar products, thus forming, to some extent, a monopoly advantage over non-price forms of similar goods. It reveals the differences between similar products and shows that they cannot be mass-produced (Chen, 2009). Housing is viewed as such a heterogeneous commodity in terms of structural, neighborhood and locational characteristics, so the Hedonic Price Model is applicable for analyzing the housing price mechanism (K. W. Chau and Chin, 2002).

Another supporting assumption is that the implicit nature of the market. The “implicit nature” means the prices and transactions of heterogeneous commodities are observable within the production, exchange and consumption process, but the price corresponding to the attribute characteristics of each commodity cannot be directly observed, which is the invisible market (Rong, 2014). The implicit market is difficult to analyze on the basis of traditional economic models because it is composed of the prices corresponding to bundles of attributes of commodities. However, the Hedonic Price Model provides a way to determine the several implicit prices of attributes (Dunse and Jones, 1998).

In spite of these debatable assumptions, which encompass considerable simplification and abstraction from a intertwined reality, the model has already been applied widely in the studies of residential housing (Chau, Ma, et al., 2001, Freeman, 1981, Leggett and Bockstael, 2000). As what has been astutely captured by Freeman (1981, p.155), “the data may be inadequate; variables are measured with error; and the definitions of empirical variables are seldom precise, but these do not render the technique invalid for empirical purposes”.

2.3 The Application of Hedonic Price Model

As what has been illustrated above, unlike other consumption goods, residential property is a kind of commodity with high heterogeneity. Even at one specific time point, prices vary greatly depending on their unique set of structural characteristics, neighborhood characteristics, and locational characteristics of the houses (Tse and Love, 2000). According to the paper of Malpezzi (2003), in which he portrayed a brilliant review of the historical development of the Hedonic Price Model, the empirical method utilized to value the effect of different components on houses is the Hedonic Price Model as it allowed the total transaction price to be decomposed into the values of various determinants. In order to explore the impact of various determinants on housing prices, the Hedonic Price Model has been widely applied into the empirical research in either foreign housing markets or Chinese housing markets over the past 40 years.

From a practical perspective, the choice of dependent variable is very imperative. Housing price usually refers to the transaction price, standing as a proxy to measure the value of one house (Sirmans, et al., 2005). As for the independent variables, there are literally hundreds of indicators under common usage. And, typically, the several determinants could be summarized into three aspects: structural characteristics, neighborhood characteristics and locational characteristics (Goodman, 1978, Williams, 1991).

2.3.1 Structural characteristics

Structural characteristics describe the condition of the physical structure of the housing. Researchers have pointed out that if one property has more desirable characteristics than other houses, all these characteristics will result in higher housing price (Ball, 1973). As Follain and Jimenez (1985) illustrated, the attributes with the highest frequency of use from previous studies were lot size, floor area, number of rooms, bedrooms and bathrooms, which have positive effect on housing price. Also, the structural condition is usually measured by age, garage, heating system and so on.

Countless research have marked out that with the increase of the numbers of rooms, bedrooms (M. M. Li and Brown, 1980) and bathrooms (Gillard, 1981) and the lot size (Rodriguez and Sirmans, 1994) and the floor area (Carroll, Clauretie, et al., 1996), the total transaction price of house will become higher. Consumers have willingness to pay for more space, especially functional space, so as to improve the quality of life (K. W. Chau and Chin, 2002) For the structural quality part, the sign of age may depend on the study period or the target city, but usually it shows a negative sign (Xiao, 2017). In those historic cities, age exerts a positive effect on the housing price because of the historical significance and vintage effect. Moreover, research pointed that different coefficients of age may correspond to two aspects: "a pure cross-sectional depreciation and obsolescence component and a demand-side component that changes over time" (Clapp and Giaccotto, 1998, p.415).

Other studies claimed that attributes that appear frequently are basement, garage (Forrest, Glen, et al., 1996), water heating system, fireplace and air heating system. Because of some endemic dampness problems, the sign of basement may be negative, but it will show positive effect most of the time. Garage typically posed a positive influence on the total price on a house although it showed insignificant among several researches. Besides, according to the study of Garrod and Willis (1992), a garage added 6.9% to the price of a dwelling, and when the garage became a double one, it would increase 20% of the price. Also, residential buildings with a central heating system worth 6.5% more than those without that. The attribute of fireplace usually shows up in western countries while it is not a common furniture in Chinese families.

Generally, depends on different preference of buyers and custom environment, the significance and expected sign of these structural attributes may not always be identical among nations Kohlhase (1991). Also, the determinants above are all tangible attributes, whereas some intangible factors such as physical and environmental quality of the properties, the developers' reputation are also correlated with housing price. Buyers are very concerned about the qualifications of real estate developers, and properties developed by reputable developers are more trusted by consumers than the others (Chen, 2009).

2.3.2 Neighborhood characteristics

Commonly, neighborhood characteristics describe the quality of socio-economic condition of a neighborhood. Although it is difficult to capture their value in the housing market, they could be implicitly valued in a Hedonic Price Model. Also, as a caveat, it should be borne in minds that failure to include neighborhood characteristics into the model will result in substantial bias (Goodman, 1978, Linneman, 1980).

Based on the previous studies, neighborhood characteristics could be generally divided into three parts (K. W. Chau and Chin, 2002):

(1) Socio-economic determinants, such as social class (Richardson, Vipond, et al., 1974), ethnic composition (Wen and Jia Shenghua, 2004, Ketkar, 1992).

(2) Government or municipal public amenities, for instance, shops (Wen and Jia Shenghua, 2004), schools (Clark and Herrin, 2000, Clauretie and Neill, 2000) and hospitals (Huh and Kwak, 1997).

(3) External influences, basically mean natural environment public spaces (Wen, Li, et al., 2012), crime rate (M. M. Li and Brown, 1980), traffic noise (Palmquist, 1992), airport noise (Espey and Lopez, 2000) and so forth.

From the perspective of social-economic factors, in one article, Richardson (1974) pointed out that the composition of social class exerted an influence on housing price. Also, some research, targeted in the United States, showed that ethnic composition was one of the significant factors among socio-economic determinants (Ketkar, 1992).

With respect to the municipal public amenities, as shops could provide convenience for people and meet their daily demand, they usually posed a positive effect on dwelling price (Wen and Jia Shenghua, 2004). Scholars generally believed that the quality of the school would impose a positive effect on the prices of surrounding homes(Clark and Herrin, 2000). In terms of hospitals, some researchers used the Hedonic Price Model to study the relationship between medical facilities as scarce resources and the surrounding house prices. The results indicated that, within 0.713km, housing prices went up as the weighted distance from the hospital to the hospital increased. Between 0.713 and 2.048 km, housing prices decreased with the increasing of the weighted distance. Outside the range of 2.048 km, housing prices increased with the upward trend of weighted distance (Si and Shi, 2013).

In terms of externalities, according to Wen (2012), she marked out the inverse relationship between distance from the public spaces or the natural environment to the sample houses and the housing price. Scholars also found that the dwelling price would go down 7.28% with one additional murder among 10,000 residents (Clark and Herrin, 2000). Besides, robbery and aggravated assault were also treated as means to measure crime rate in some other surveys (Haurin and Brasington, 1996).

2.3.3 Locational characteristics

The locational attributes are commonly qualified from the perspective of the entire city range, usually being quantified by locational accessibility and transportation convenience (Follain and Jimenez, 1985, Ming and Hou, 2017). To be more specific, locational accessibility is quantified by measuring the distance from a plot to the city center, while transportation convenience is gauged by measuring the distance from a house to the nearest bus stop or subway station (Ming and Hou, 2017).

In terms of the locational accessibility, in the Von Thunen's model, which is under the assumption of a monocentric city, the distance to CBD (Central Business District) has become a necessary determinant among most of the studies on Hedonic Price Model. Frequently, the distance from CBD usually exerts a negative impact on the housing price, as the commuting cost will increase. However, the assumption and the sign of impact have increasingly been criticized with the theoretical development and innovation (Boarnet, 1994). Since more and more cities have become polycentric with the economic development, the distance from CBD may not be the unique measurement of locational accessibility. Moreover, the shorter the distance from a residential property to a commercial center, the more the noise pollution, air pollution, and traffic jams in the urban areas will have to face. Therefore, the impact of the distance from the commercial center on the housing price is not simply monotonous. Secondly, when comes to the transportation convenience, it is usually measured by the accessibility to transportation tools (Baldassare, Knight, et al., 1979). Generally, most researchers agree on that rail transit exerts a positive impact on residential housing value. And this impact is stronger in downtown compared with suburbs (K. Liu, Wu, et al., 2015). What is more, the effect has difference between long run and short terms. In the short run, the new-built rail transit forms will lead to a suddenly rising in housing price while in the longer term, this kind of influence will remain stable (Su, Zhu, et al., 2015).

2.3.4 Summary

The literature review of previous research shows that there are several typical attributes that have been packed in most of Hedonic Price Model studies. Table 1 displays a general list of the common attributes, based on the review above. Also, since the Hedonic Price Model is derived from foreign countries, the selecting of factors is more likely to suit foreign conditions. As a result, when targeting at China, some attributes should be changed corresponding to specific Chinese situation. For instance, there are no fireplaces and basements in typical Chinese houses.

Table 1 List of commonly used of determinants of Hedonic Price Model (Source: the author based on the literature review)

Characteristics	References	Determinants	Expected impact
Structural characteristics	Li and Brown (1980)	Number of rooms, bedrooms	+
	Rodrigue and Sirmans (1994); Carrol, Clauretie et al. (1996)	Lot size and the floor area	+
	Clapp and Giaccotto (1998)	Age of the building	-/+
	Gillard (1981)	Number of bathrooms and heating system	+
	Chau and Chin (2002)	Floors and building type (multi-storey etc)	+
	Garrod and Willis (1992)	Basement, garage	+/-
	Chen (2009)	developers' reputation	+
Neighborhood characteristics	Wen Haizhen (2004)	Shops	+
	Wen, Li, et al. (2012); Wen Haizhen (2004)	Accessibility of the nearby green land and public area; Accessibility of the nearby water system	+
	Clark and Herrin (2000)	Accessibility of the nearby schools, especially primary schools and middle schools	+
	Huh and Kwak (1997)	Hospitals	+
	Clark and Herrin (2000, 1996); Haurin and Brasington (1996)	Crime rates	-
	Williams (1991); Espey and Lopez (2000)	Traffic noise; Airport noise	-

Locational characteristics	Grass(1992) and Ryan(2005)	Locational accessibility (distance to CBD)	-
	Simons R.A. and Jaouhari(2004)	Transportation convenience (bus or rail transportation)	+

+ means positive effect on housing price. – means negative effect on housing price.

2.4 Spatial Heterogeneity in Attributes of Housing Prices

2.4.1 Spatial heterogeneity

With the development of the spatial analysis, interest about applying spatial heterogeneity of Hedonic Price Model has grown rapidly among researchers in a global context (Helbich, Brunauer, et al., 2013, Lu, Charlton, et al., 2014). Researchers are compelled by the notion that systematic variation in housing prices in different spatial locations, resulting in heteroskedastic error terms. For example, as the distance to the city center increases, the price of housing continues to decline, but the amplitude of price fluctuations varies significantly. Scholars usually explain the spatial heterogeneity of housing price based on Hedonic Price Model or from the perspective of residential submarkets. Because there are obvious differences among various features that make up the value of residential use, such as the location, floor, area, interior design, and residential environment, the house itself is a typical heterogeneous product. Especially, scholars pointed out that the differences in residential properties and price fluctuations due to different locations are the basic characteristics of the urban residential market (Straszheim, 1975). Another explanation for the spatial heterogeneity of residential housing prices is the residential submarket theory. The theory holds that due to the constraints of some physical barriers within the housing market, the entire housing market in one area has already been divided into residential submarkets, and the structural differences within the submarkets have led to spatial heterogeneity of residential prices in different submarkets (Osland, 2010).

From the perspective of statistics, there is inevitably some spatial heterogeneity in the Hedonic Price Model because of the above reason. However, the Hedonic Price Model, based on Ordinary Least-Squares (OLS) regression, could only analyze the linear relationship between the housing attributes and the house price differentials without giving adequate attention to spatial heterogeneity (Paez, Uchida, et al., 2001). That is to say, a hypothesis is implied in the study that the effect of the independent variables is spatially uniformed. This assumption is obviously not completely consistent with reality or is an unrealistic simplification of residential dwelling markets (McMillen and Redfearn, 2010). However, the results estimated by utilizing the traditional Hedonic Price Model which neglects the spatial variation will lead to biased coefficients, a loss of explanatory power and inaccurate conclusions (LeSage and Pace, 2004, X. Liu, 2017). When it is applied into designing policy strategies, policymakers have to take the risk of identifying incorrect submarkets and building erroneous housing assessments (Helbich and Griffith, 2016). However, according to the research of Bitter et al. (2007), the previous studies based on OLS estimations are basically collapsed in that they fail to adequately consider and estimate spatial heterogeneity. Consequently, it is critical for researchers to consider spatial heterogeneity and build correct spatial econometric models that have the highest fitness (Bourassa, Cantoni, et al., 2010, Ahn, Byun, et al., 2012).

2.4.2 Spatial econometric estimation techniques

The general spatial econometric estimation techniques to deal with spatial heterogeneity within housing markets are Spatial Expansion Method (SEM), Moving Window Regression (MWR) and Geographically Weighted Regression (GWR).

SEM, the precursor of GWR estimation, is a classic method to explore the spatial instability of housing prices (Osland, 2010). However, its polynomial expansion is criticized to be too inaccurate to estimate spatial heterogeneity effectively (Pace, Barry, et al., 1998). Also, because there will be numerous interaction terms in a SEM, researchers have to limit the number of variables, which may result in omitted bias (Fik, Ling, et al., 2003). To solve this problem, MWR and GWR estimations are invented to generate distance-based weightings according to the coordinate of each sample housing. As spatial heterogeneity in coefficients may be considered to exist in a continuous or discontinuous way, the GWR estimation and MWR estimation are used to test for spatial parameter heterogeneity (LeSage and Pace, 2004). Since every estimator has its own limitations, both of GWR and MWR are sensitive to multi-linear problems. They assume there are linear relations between two variables even if they actually have no linearities (Helbich and Jokar Arsanjani, 2015). Farber and Yeates (2006) also tested GWR estimation against MWR estimation method and reported that GWR estimation was more accurate and had more explanatory power. For the theoretical and empirical perspectives, GWR estimation seems to be better, although it has shortcomings. Besides, as GWR estimation has already been implemented in its independent software, it is more 'user-friendly' than other estimation techniques (Brunsdon, Fotheringham, et al., 2002).

2.5 Literature Commentary

Residential property is a highly heterogeneous commodity, also, these heterogeneous factors are valued discrepantly by different consumers (Sirmans, et al., 2005). Even at the same time, prices vary greatly depending on the structural, neighborhood and locational characteristics. Researchers in the global context mainly adopt the Hedonic Price Model to reveal the influence of residential characteristics on house prices. The Hedonic Price Model is not only statistical, but also with a theoretical framework derived from consumer theory (K. W. Chau and Chin, 2002). To realize its theoretical foundation, the Hedonic Price Model uses ordinary least-squares method to reflect the average impact of housing attributes on the total price without considering the spatial heterogeneity. With the deepening of research, scholars proposed a new method of spatial analysis—a Geographically Weighted Regression estimation. This GWR method can effectively reveal the spatial dependence among variables within the observed space. The application of geographically weighted regression estimation technique is broadly in many fields, especially in studies of the impact of the spatial factors exerting on housing prices. The geographically weighted regression is decomposed into several spatial parameters estimate which is better than the overall average parameter estimation provided by the least squares method. It can profoundly estimate the complex relationship between residential housing prices and spatial influence factors, and the visual tools can also give more details in the form of geo-maps (J. Liu, 2016).

2.6 Conceptual Framework

The conceptual framework of this research, as is shown in Figure 3, embeds spatial heterogeneity into the analysis of the relationship between housing attributes and housing price. After deciding to target at the second-hand houses, determinants of housing price are identified based on the literature review. According to the literature view, the Hedonic Price Model, served as a general method, are introduced to detect the housing characteristics. Namely, the structural characteristics, neighborhood characteristics and locational characteristics. In the next step, an OLS estimation is used to analyze the basic correlation between housing characteristic and housing price. After that, a Moran's Index will be calculated to test if there is spatial autocorrelation problem in the OLS estimation, which will cause the inaccuracy of the estimation. The last step is to build a GWR estimation with a higher fitness and accuracy to incorporate the spatial heterogeneity in the determinants of second-hand housing price.

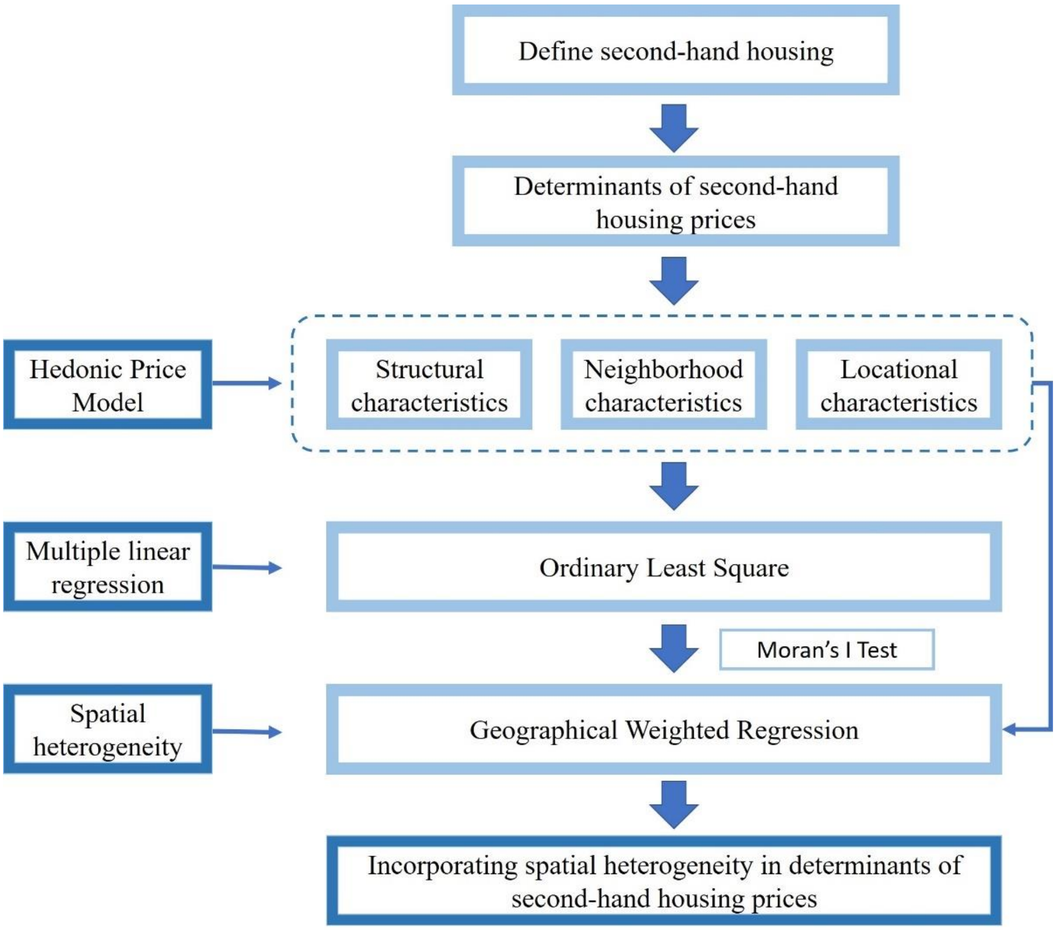

Figure 3 The conceptual framework (Source: the author)

Chapter 3: Research Design and Methods

3.1 Research Question

The main objective of this research is to explore what determinants and how they will affect the price of second-hand houses in Zhengzhou City, also to examine the spatial heterogeneity in determinants of second-hand housing price in Zhengzhou City. After the above literature review part, the research can be recapped as follows:
The main question is:

What are the determinants of second-hand housing prices and do these determinants vary between different areas?

The sub-questions are:

a. Based on Hedonic Price Model, what factors and how they will influence the second-hand housing prices in Zhengzhou City?

b. Is there spatial heterogeneity in the price composition of the second-hand housing market in Zhengzhou City?

3.2 Operationalization

As what has been discussed in the literature review, there are two main parts in this research: determinants of second-hand housing prices and spatial heterogeneity. The first part, imbedded in the first sub-question can be analysed in three aspects. Respectively, they are the structural characteristics, the neighborhood characteristics and the locational characteristics. Also, based on the traditional Hedonic Price Model and the practical situation in China, the variables and indicators that will influence the second-housing prices in Zhengzhou City are listed in Table 2. While taking the spatial heterogeneity into account, the basic determinants of second-hand housing price remain the same, only need to be put in the Geographically Weighted Regression estimation method.

For the dependent variable, the housing price will be utilized in this study which refers to the transaction price of second-hand housing. Its natural log formation is selected as experimentation showed that the semi-log function performed better than other functional forms (Sirmans, et al., 2005). As regarding the independent variables, this part will be composed of variables and indicators of structural, neighborhood and locational characteristics.

3.2.1 Structural characteristics

In this part, there are two imperative concepts imbedded in structural characteristics, respectively the functional space and functional period. To measure the functional space, building size according to the record on the property ownership certificate is firstly used as space has relation with the quality of living. Also, floor is significant to a house. Taking low-rise housing as an example, it is popular to say in China that "the third floor is gold and the fourth floor is silver" which means the third or fourth floor are the best choice of an ordinary low-rise building. For high-rise buildings, the higher the better, because houses on the higher floors have better sunshine and broaden fields of view. The building type including low-rise building and high-rise building which will show the impact of people’s different preference of different functional buildings on housing price. Moreover, the orientation of the house will affect the lighting and ventilation, in this light, deciding the value of a house. On one hand, people traditionally prefer the housing facing south influenced by Chinese culture, on the other hand, China is located in the northern hemisphere and houses in the south have good daylight. Except the considering of functional space, the functional period measured by the indicator "building age" is extremely important to a single house. To be specific, the later it is built, the more years it can be used as the life expectancy of a residential house is 70 years according to the law. Also, their construction style will be updated to people's taste.

3.2.2 Neighborhood characteristics

Neighborhood characteristics mainly concern the environmental livability and accessibility of public services. The environment can be decomposed into residential environment and natural environment. As for the residential environment, the plot ratio was selected. The higher the floor area ratio usually indicates that the building is higher and the area of the open space per person is smaller. This kind of neighborhood gives people a sense of oppression, and an excessively high plot rate will inevitably affect the distance between buildings and affect the lighting further. The natural environment livability is composed of the inner green land and that outside the neighborhood. To be exact, the higher green ratio of the neighborhood not only helps to purify the air, but also adjusts the temperature difference in the plot, and it also brings people a good mood. Besides, large parks or green spaces will be set as a symbol of outside environment measured by numbers of parks within 1 km in this research. Parks can not only provide good public space for the community, but also adjust the local climate and improve the overall score of the neighborhood. The price of residential houses near parks or green areas tends to be high, which is also an important factor affecting property prices. While the accessibility of nearby public services, mainly focusing on the hospital, primary and middle school, reveals the functional role of one house. From the empirical perspective, a house with a better accessibility to a hospital can be sold with a higher price since people concern the healthy issues. However, hospitals sometime decrease the price of one house because of the noise of ambulances or other traditional Chinese taboos. Also, the equality of education package is particularly important for the impact of real estate. Parents hope to enjoy educational resources with good quality in places close to their homes. Since there is a school district housing system in Zhengzhou City, it is common to increase the willingness of parents to pay a higher price for the houses in the district of a Grade-A primary school.

3.2.3 Locational characteristics

The public transportation and distance to CBD also measure the accessibility of public services of a house. As for a suitable residential area, the convenience of transportation is particularly important, the more convenient the traffic, the greater the impact on the value of residential real estate. The traffic convenience of one house in this study is judged by the number of bus stops within 1 km, the distance to the nearest subway station and the distance to the nearest arterial road. According to the ground tax theory, land value reduces concentrically with the commercial area as the center. Corresponding to the context of Zhengzhou City, the property price is expected to show a downward spiral with the increase of the distance to Erqi Square where the CBD locates.

Except for the indicator 'Distance to Jinshui River Park', all of the other indicators are mainly sourced from the literature review. Some researchers indicate that large-scale urban landscape has positive impact on housing prices (Wen, et al., 2012). In the past 2,500 years, the Jinshui River was like a streamer, flowing from west to east, flowing around Zhengzhou citizens and forming a close relationship with the city of Zhengzhou. In this study, Jinshui River Park will be considered as a traditional large-scale land scape and its effect on second-hand housing prices in Zhengzhou City will be measured in the next Chapter.

Table 2 Operationalization and methodology

Characteristics	Concepts	Variables	Indicators	Data Type	Data source	Expected impact
Housing prices	House value	Second-hand housing prices	Second-hand housing transaction price in RMB	Quantitative secondary data	Official data from the Henan Local Taxation Bureau	N/A
Structural characteristics	Functional space	Building size	Building size in square meter			+
		Floors	Which floor is the housing on			+/-
		Building type	Dummy variable: low-rise building=0 (<28 storeys), high-rise building=1(>=28 storeys)			+/-
		Building orientation	Dummy variable: South=1, North=0			+
	Functional period	Building age	Building age			-
Neighborhood characteristics	Environmental livability	Plot ratio	Plot ratio of the neighborhood		Real estate agency data from Anjuke	-
		Green ratio	Green ratio of the neighborhood			+
		Accessibility of the nearby green land	Distance to the nearest park		Electronic map data from API and ArcGIS	+
	Accessibility of public services	Accessibility of the nearby hospital	Distance to the nearest hospital			+/-
		Accessibility of the nearby education	Dummy variable: if it is in the first-class primary school district. Yes=1, No=0			+ +
Locational characteristics		Transportation convenience	Number of bus stops within 1 km		Real estate agency data from Anjuke	+
			Distance to the nearest subway station			-
			Distance to the nearest arterial road		Electronic map data from API and ArcGIS	-
		Locational accessibility	Distance to Erqi Square(CBD)			-
			Distance to Jinshui River Park			-

+ means positive effect on housing price, – means negative effect on housing price.

3.3 Data Collection Methods

The quantitative secondary data collection will be presented as the main data collection method in this research. To obtain the information of the above indicators, data are derived from the following three methods.

3.3.1 Official data

First, obtaining relevant data on second-hand housing transaction prices and structural characteristics of each sample house is one of the main quantitative data collection method in this study. The secondary data on the details of all the housing transactions in 2016 will be obtained from the Henan Local Taxation Bureau, including the transaction prices, building size, floors, building orientation, building type and building year.

3.3.2 Electronic map data

Also, in order to collect data on neighborhood characteristics and locational characteristics, the Baidu map coordinate picking system (API)[3] and the software ArcGIS will be used in this part. Firstly, by using the Baidu map coordinate picking system (API), the coordinates of all the second-hand sample houses will be obtained. Moreover, based on ArcGIS (GCS China Geodetic Coordinate System 2000), the electronic database of urban facility shapefiles in Zhengzhou City, including the district map, the distribution of hospitals, primary schools, middle schools and parks, and the second-hand sample houses. According to the formal document of primary and secondary education service area announced by the Zhengzhou Municipal Education Bureau in 2015, ArcGIS will determine whether the sample house point

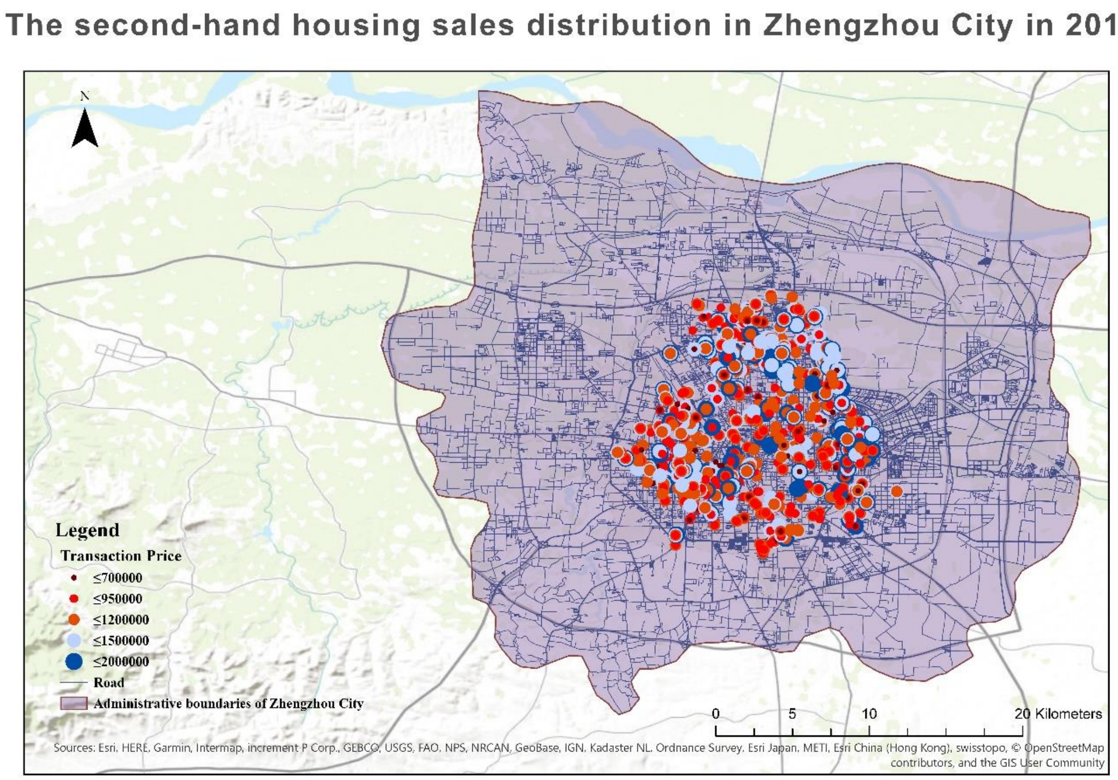

Figure 4 The distribution of housing samples (Source: the author)

is located in the education district. In the same way, the accessibility of the nearby green land

[3] http://api.map.baidu.com/lbsapi/getpoint/index.html

and hospitals can also be measured through ArcGIS. The spatial distribution of all the second-hand housing sample points studied in this thesis is shown in Figure 4 and all these points are classified by their transaction prices in RMB.

3.3.3 Real estate agency data

The other types of data in this research are derived from a well-known real estate trading agency in China, Anjuke[4] . Established in January 2007, Anjuke is the No. 1 real estate rental service platform in China, focusing on real estate rental and sales information services. According to the evaluation results, Anjuke becomes the most comprehensive real estate APP with its functional products and higher market coverage. Till now, the Anjuke website has over 69 million visitors every month and its App users have exceeded 25 million, accounting for 70% of the market coverage for renting or buying houses by mobile terminals. In this study, some of the neighborhood characteristics and locational characteristics, for example, the green ratio and plot ratio, number of bus stops and subway stations within 1km will be attained through the Anjuke website.

3.4 Sample Size and Selection

As shown in Figure 5, Zhengzhou City, located on the southern bank of the Yellow River and occupied the kernel part of China, is the provincial capital of Henan Province. The research

Figure 5 Location of Zhengzhou City in China

area is the main urban area of Zhengzhou City, excluding the Shangjie district which is an exclave in anther district in Henan Province (Figure 6). This is because Shangjie district is separated by space and its real estate market is quite different from the main area of Zhengzhou City. After considering the data availability, the sample size for both the

---

[4] www.anjuke.com

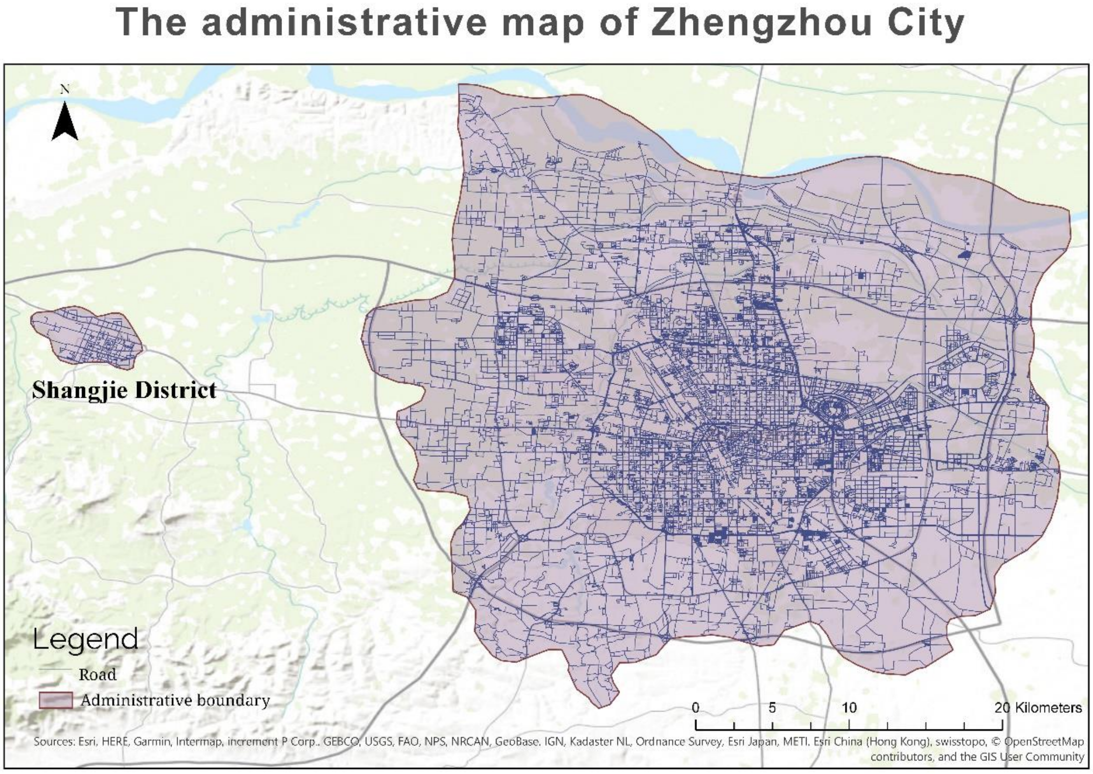

Figure 6 The administrative map of Zhengzhou City (Source: the author)

dependent indicators and independent indicators is based on the real transaction situation in Zhengzhou City in 2016. The total number of residential transaction housing is 12684, and after deleting the sample house in Shangjie district, the sample size is 10524. Considering the processing power of the GWR 4.0, which is the software to build the GWR estimation, the final sample will be sampled randomly from the 10524 and the number is 6684.

3.5 Validity and Reliability

3.5.1 Validity

Validity measures whether a study correctly reflects or assesses the concepts and it contains internal validity and external validity. Regarding to the internal validity, the variables and indicator were selected based on the literature review which aimed at measuring the corresponding concepts. While the external validity asks for that if the research could be generalized to other similar situations. Generally, the indicators were typical determinants of second-hand housing prices and can be applied in other study areas. However, only targeted in Zhengzhou City, this study chose the special landscape "Jinshui River Park" as an experimental indicator to explore the value of landscape effect. When analyzing the influence of determinants on housing prices in other areas, the typical local landscape should be reconsidered according to the corresponding micro context.

3.5.2 Reliability

Reliability means if repeated tests will result in the consistent results. For this research, reliability was safeguarded by the data source firstly (Healy and Perry, 2000). They are derived from local taxation bureau and other authoritative search engine and real estate agency with good domestic reputation, which will ensure the reliability of the study and increase the interpreted ability of the empirical results. Also, a cross-checking data from other real-estate agency websites was selected to guarantee the triangulation process.

3.6 Data Analysis Methods

To begin with, an Exploratory Spatial Data Analysis（ESDA）is used to grasp the basic distribution of the housing price in Zhengzhou City, which will provide a foundation for studying the correlation between the housing price and housing attributes in the next step. After that, to analyze the impact of the determinants imposing on second-hand housing prices and further to take the spatial heterogeneity into account, an OLS estimation based on the Hedonic Price Model and a GWR estimation should be adopted respectively.

3.6.1 Exploratory Spatial Data Analysis（ESDA）

Exploratory Spatial Data Analysis (ESDA) can detect spatial properties or the hidden features of data sets. As a “data driven” analysis method, it is used to delineate the spatial patterns, formulate hypotheses and explore correlations based on the essential map (Anselin, 1999). It is based on the principle of statistics, using Geographic Information System (GIS) technology to analyze and detect the distribution characteristics and spatial properties of one dataset with intuitive tables and graphs, thus laying the foundation for exploring the variables and structures of mathematical models.

Compared with the traditional statistical analysis technology, the ESDA method does not pre-set the characteristics or regular pattern of the data, but gradually analyses the data, and discovers the regularity embedded in the data. To be specific, a probability distribution analysis, a spatial trend analysis and a spatial autocorrelation and agglomeration analysis are introduced to analyse the spatial patterns and characteristics of second-hand housing prices in Zhengzhou City. The first two analyses will explore the spatial trends and distribution patterns of the housing prices in Zhengzhou City. While the spatial autocorrelation and agglomeration analysis will go deeper to identify the global concentration and local correlations of the housing price.

As Arbia (2001) points out, an appropriate method should capture two dimensions of agglomeration: spatial concentration within any one spatial unit and the spatial correlation among all the units across the discrete study area. More precisely, the method should gauge the degree of spatial agglomeration and reveal the correlation with other units in this area (Guillain and Le Gallo, 2010). Since the overall geographical concentration and the local correlation tend to be analysed separately, the Global Moran’s I and Local Moran’s I are introduced to present the two different patterns.

3.6.1.1 Spatial autocorrelation analysis (Global Moran’s Index)

The Global Moran’s Index is the most widespread used method to measure the autocorrelation between location attribute values. It is a translation of numerical data to a spatial correlation which is usually applied to areal units (Moran, 1950). The corresponding I is calculated by the formula:

$$\mathrm{I} = \left[\frac{n}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right] * \left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij(y_i-\bar{y})(y_j-\bar{y})}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\right]$$

In this formula, y represents the attribute value of each region, and the subscripts i and j represent different regions. $\bar{y}$ indicates the mean value of the population and n refers to the number of regions. w_{ij} represents the spatial weight of i and j. It can also be translated in a matrix form:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} * \frac{y^T W_y}{y^T y}$$

The spatial autocorrelation index can reflect the spatial aggregation distribution features of housing price. By measuring the Moran's Index, the spatial correlation degree and degree of difference of the whole region can be analysed. The index ranges from -1 to +1. If it is positive, the spatial data is positively correlated and distributed aggregately. A negative value shows a negative correlation and the points distributed discretely. A value of 0 means there is no autocorrelation. A value of 0.3 or greater, or a value of -0.3 or less, indicates a relatively strong autocorrelation (O'sullivan and Unwin, 2014).

The Spatial Autocorrelation (Global Moran's I) tool in ArcGIS calculates spatial autocorrelation based on the coordinate and value of features. Given a set of characteristics and related attributes, the tool identifies whether the expressed pattern is a clustering mode, a discrete mode, or a random mode. The significance of the index is usually evaluated by calculating the z-score and p-value.

3.6.1.2 Local spatial autocorrelation analysis (Local Moran’s Index)

The local spatial autocorrelation analysis mainly analyses the distribution pattern of each attribute value in the heterogeneous region and measures the degree of local correlation between each position attribute value and the neighbouring area attribute value. Anselin (1999) proposed local indicators of spatial association (LISA) and Moran scatterplot, which identified the spatial clustering and spatial outliers of the data. To be specific, LISA statistics and Moran scatterplot pinpoint a four-way split of the sample, including the high-high cluster, high-low cluster, low-high cluster and the low-low cluster. The equation of Local Moran’s Index can be written as following:

$$I_i = \sum_{j=1}^{n} W_{ij}(X_i - \bar{X})\,(X_j - \bar{X})$$

In this formula, n indicates the number of the subjects, X_i and X_j are the values of observations, while W_{ij} refers to the spatial weighted matrix. After getting the value of the Local Moran’s Index, it is also necessary to judge the significance based on the Z-score, which is also the same as that in the global spatial autocorrelation analysis. The positive value of Local Moran’s I represents that the variables with similar values have a spatial aggregation. Whereas, the negative Local Moran’s Index means a feature has neighbouring points with dissimilar values. Therefore, there are four correlation modes in the sample: when I_i>0 and Z>0, it shows a local high-value spatial clustering (spatial connection form of high-high); when I_i<0 and Z>0, it indicates a local low-value clustering (spatial connection form of low-low); when I_i<0 and Z<0, it is a local high-valued heterogeneous type (spatial connection form of high-low); When I_i>0 and Z<0, it is a local low-valued heterogeneous type (spatial connection form of low-high) (Guillain and Le Gallo, 2010).

3.6.2 OLS estimation technique based on the Hedonic Price Model

As outlined in the article of Fotheringham et al (2000), the classical Hedonic Price Model formula which could illustrate the impact of a bundle of determinants on housing price can be shown as follows:

$$y_i = \alpha_0 + \sum \alpha_k \, x_{ik} + \varepsilon_i$$

In this formula, y_i represents the ith observation of the housing price, α_0 means the constant term, x_{ik} is the kth characteristic variable, α_k refers to the parameter of the corresponding characteristic variable that will be estimated, and ε_i is the error term.

However, before operating the OLS regression, there are six necessary OLS assumptions:

Assumption 1: Linearity: A linear relationship between Y and X variables is required.

Assumption 2: Outliers: No significant outliers, as they have over-influential residuals.

Assumption 3: Multicollinearity: Independent variables not to correlate too much, or not to be a linear function of each other.

Assumption 4: Homoskedasticity: The variance in the error terms must be homoscedastic.

Assumption 5: Normality: The error terms need to be normally distributed.

Assumption 6: Omitted variables: Check for possibility of omitted variables.

3.6.3 GWR estimation technique

The Geographically Weighted Regression (GWR) shows extremely high application value and estimation accuracy in the research practice of revealing the spatial heterogeneity of housing prices and the influence of each housing characteristic on prices (She, 2017). Compared with the Hedonic Price Model, GWR estimation assumes that the coefficient is a function of both the spatial position of variables and each kind of housing determinant. Similarly, the general formula of GWR estimation is:

$$y_i = \alpha_0 + \sum \alpha_k\,(u_i, v_i)x_{ik} + \varepsilon_i$$

Here (u_i, v_i) represents the spatial coordinate of the ith observation.

As the α_k changes with coordinates, each α_k is used to estimate its neighbouring spatial observations. When correcting GWR, the weight matrix is established for each location observation i, instead of directly using the nearest observation value. At its simplest, $\alpha_k(u_i, v_i)$ can be estimated as follows:

$$\alpha_k(u_i, v_i) = (X^T W(u_i, v_i)X)^{-1} X^T W(u_i, v_i)y$$

$W(u_i, v_i)$ means a spatial weighting matrix. Also, a Gaussian function is used to measure the weigh as shown in the following formula.

$$W(u_i, v_i) = exp(-d/h)^2$$

d refers to the distance between the observation point and the regression point, while h signifies the bandwidth. The value of the bandwidth sensitively affects the spatial variation in the GWR estimation and can be determined by the cross-validation method (Bitter, et al., 2007).

For the inspection part, the common method to test the significance of a GWR estimation is the Akaike Information Criterion (AIC) criterion. This method is based on the concept of entropy and is used to compare the goodness-of-fit between multiple regression estimations. The basic formula is:

$$\mathrm{AIC}_C = \ln\left(\frac{RSS}{n}\right) + \frac{n+k}{n-k-2}$$

Where k indicates the number of variables, n refers to the sample size, and RSS is the sum of the squares of the residuals. AICc is a relative dimension. For different estimations with the same independent variable, the smaller the AICc value is, the better the goodness-of-fit of the estimation. When the difference of AICc values between estimations is less than 3, the goodness-of-fit between the estimations is equal.

Chapter 4: Research Findings

In this chapter, the sample data of housing prices are exployed for analysis to explore its distribution and some hidden features. Moreover, econometric estimations and necessary tests are conducted to explore the correlation between housing attributes and housing price. Finally, some findings and conclusions are summarized to answer the main and sub questions listed in Chapter 3.

4.1 Exploratory Spatial Data Analysis（ESDA）

4.1.1 Probability distribution analysis

Based on Stata 14.2, an effective graphical histogram was constructed to show the distribution of the second-hand housing prices in Zhengzhou City (Fig.7). Also, it reveals several important statistical indicators with an overall sample size of 6684. The mean and median reflect the concentration of the sample trend. The standard deviation is used to pinpoint the dispersion of the sample data. The skewness and kurtosis can measure the distribution pattern of a data set. For a standard normal distribution, its skewness should be zero and the kurtosis should be equal to three. In this case, the median (930000) is smaller the mean (981840), which indicates that there are more smaller housing prices than those in the higher side. Also, based on a positive skewness of 1.3636, we can say that the distribution of the transaction housing prices is skewed right, which also means the right tail is long compared with the left tail. With a skewness of 0.9766 and a kurtosis of 4.2535, a conclusion can be drawn that the identity form of the housing price is not distributed normally. Therefore, it is necessary to properly convert the raw data to make it closer to the normal distribution.

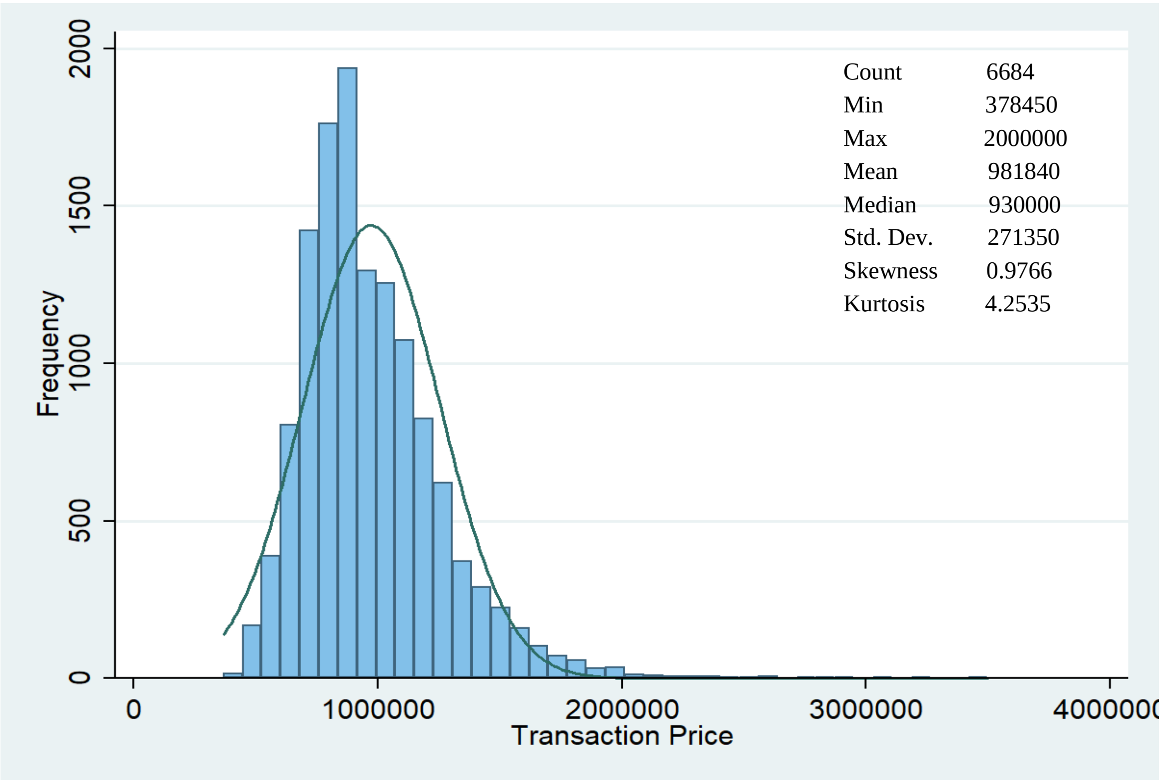

Figure 7 Histogram of transaction housing price in Zhengzhou City

As a common statistical way, the variable of transaction housing price was winsorized at its highest 31 observations to make a reduction in the impact of outliers and data errors (Cox,

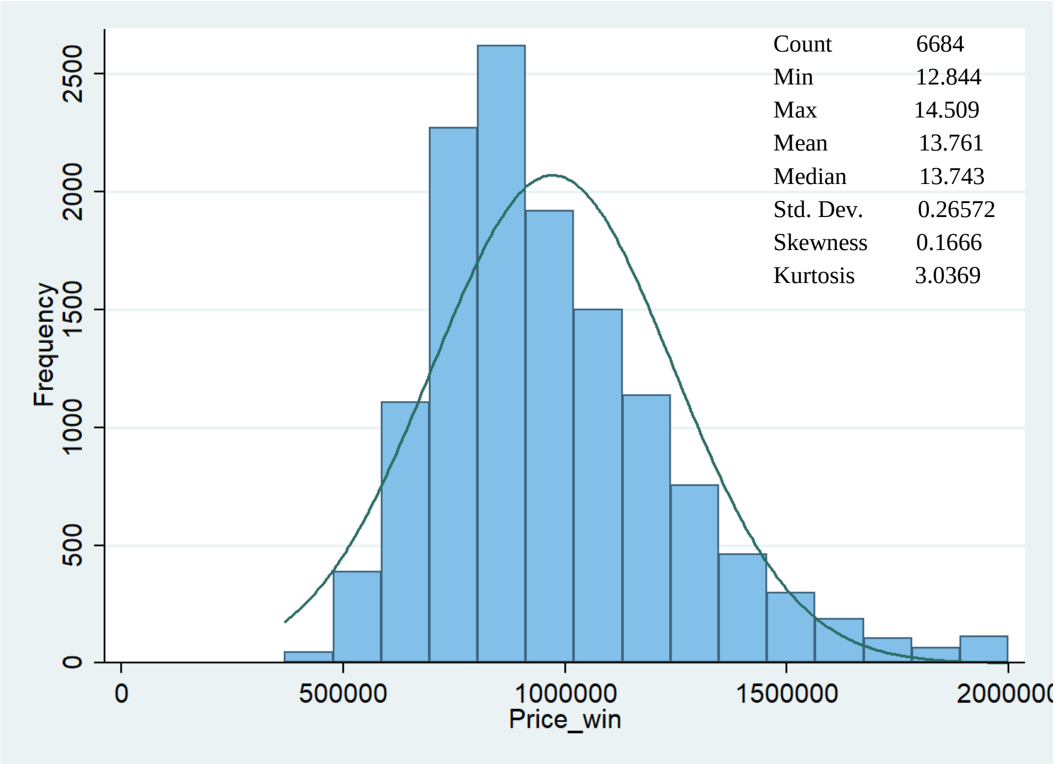

Figure 8 Histogram of logarithmic functional form of winsorized transaction housing price in Zhengzhou City

2006). To be specific, these 31 observations were replaced by the 30th largest value (2,000,000). Figure 8 shows a histogram of housing price distribution generated by natural logarithmic transformation of winsorized residential price data in Zhengzhou City. It can be seen that the housing price, after the logarithmic transformation, presents a bell-shaped distribution, and the average and median of the logarithm of the house price are closer, indicating that the data is almost evenly distributed. Besides, the skewness of the logarithmic housing prices is 0.1666 with a kurtosis of 3.0369, very closing to 0 and 3 respectively, which indicates that the sample data is basically normal distributed.

4.1.2 Spatial trend analysis

As what has been illustrated above, unlike other consumption goods, residential property has high heterogeneity. Even at one specific time point, prices vary greatly in different distributions depending on their unique set of structural characteristics, neighborhood characteristics, and locational characteristics of the houses. On the foundation of the spatial trend analysis of the ArcGIS, we can explore the spatial trends and distribution patterns of housing prices in Zhengzhou City (X. Wang, 2015).

Overall, the trend analysis can provide a sideway view of the dataset. Each data point is projected into the three-dimensional coordinate system by the trend analysis tool in ArcGIS. The longitude direction and the latitude direction are plotted on the x and y plane respectively and the housing price is measured by the z-dimension (Zhang and Sun, 2016). Figure 8, 9 and 10 are global trend analysis diagrams of the housing price in Zhengzhou City. The X axis represents the longitude, the Y axis describes the latitude and the Z axis reveals the numerical value of the transaction housing price. The light red dots on the horizontal plane indicate the position of the sample houses and the height of the vertical lines correspond to the housing price. The green dots on the XZ plane and the blue dots on the YZ plane represent the projection of the sample point values on the two planes, so as to simulate the trend in that direction which is expressed by the two curves. If the fitted line is flat, it means there is no trend within the data.

Figure 9 and 10 present the overall transaction housing price in Zhengzhou City has an inverted bell-shaped curve on both the north-south direction and the east-west direction, gradually increasing to the highest value, and then gradually decreasing. The housing prices in the western part of the district are relatively lower than those in the eastward area, while the central part has the highest housing prices. The inverted U-shaped curve in the east-west direction is not as obvious as that in the north-south direction because the rise of the Zhengdong New District has boosted the a rapid increment of the nearby housing price. The situation is similar to that of the north-south direction. Compared with the central part, the housing price in the periphery of Zhengzhou City is lower. After reaching the highest point in the middle, the curve fell slightly in the south side, but it is still highter than that of the northern part.

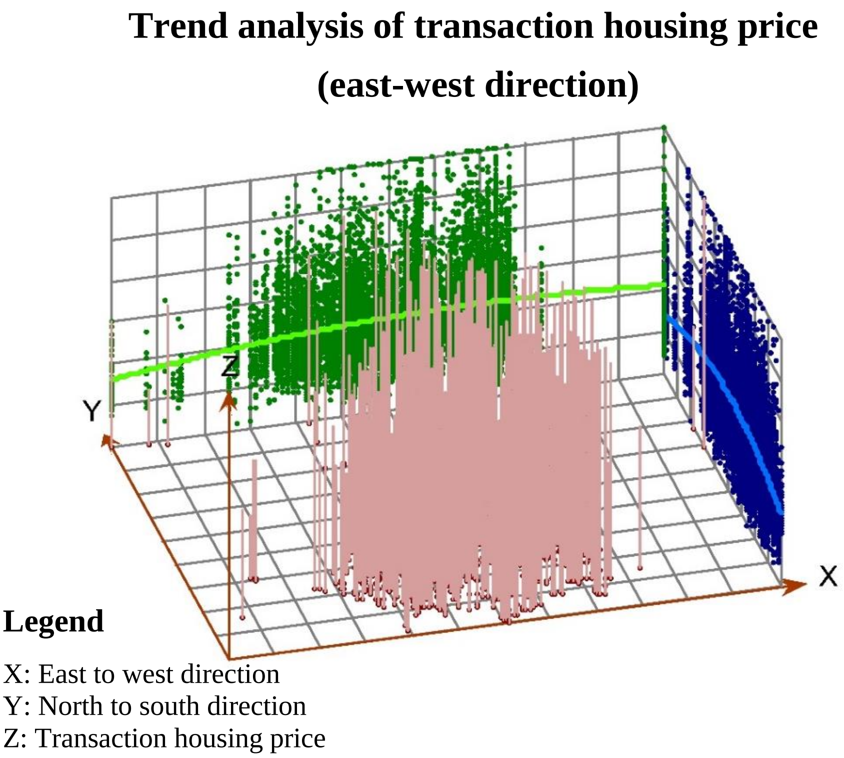

Figure 9 Trend analysis of transaction housing price (east-west direction)

Figure 11 is obtained by rotating the original image horizontally to 45 degrees. We can also find that the trend of high price in the middle and lower price in the periphery remains in the southeast-northwest direction and the northeast-southwest direction. This trend is basically consistent with the real situation of the real estae market in Zhengzhou City.

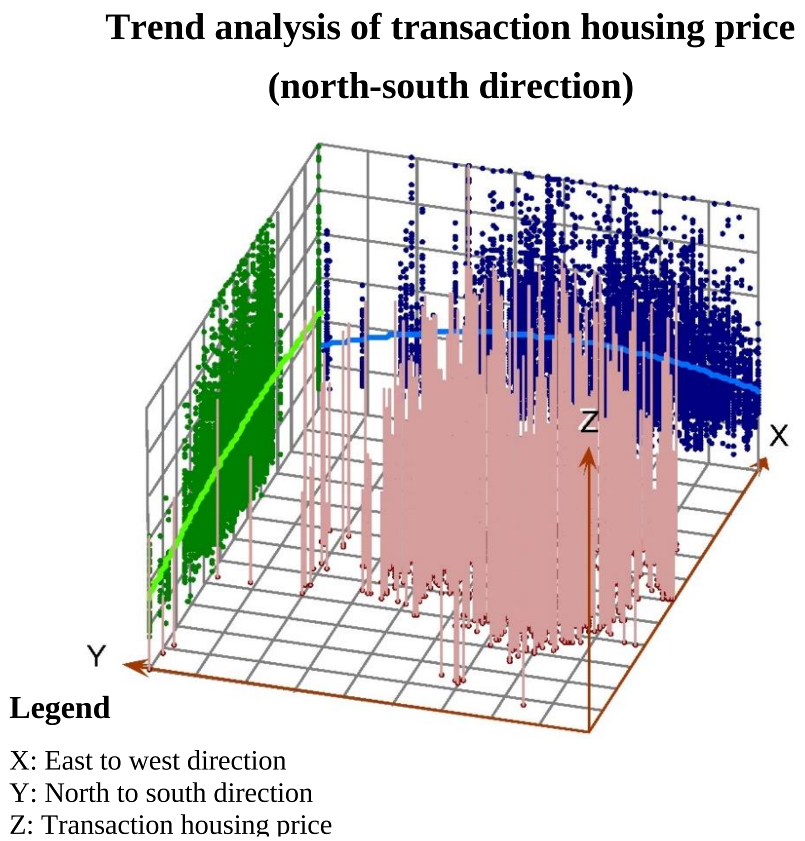

Figure 10 Trend analysis of transaction housing price (north-south direction)

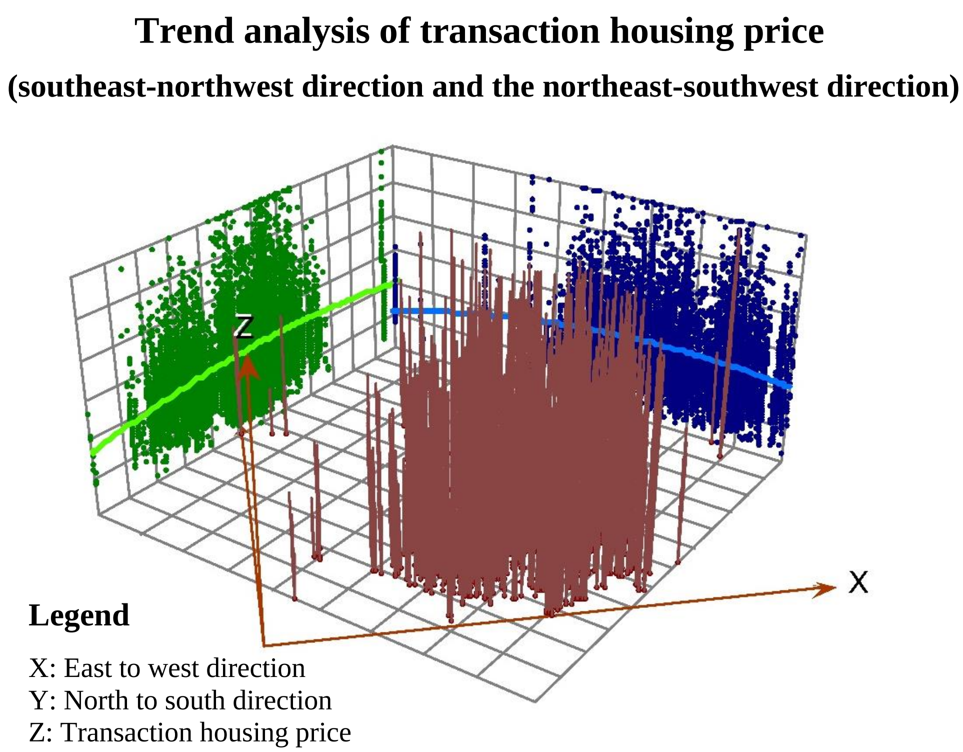

Figure 11 Trend analysis of transaction housing price (southeast-northwest direction and the northeast-southwest direction)

4.1.3 Spatial autocorrelation and agglomeration analysis

4.1.3.1 Spatial autocorrelation analysis (Global Moran's Index)

As mentioned in Section 3.6, the Spatial Autocorrelation (Global Moran's I) tool in ArcGIS is used to calculates spatial autocorrelation based on the coordinate and value of features. In this case, gauged by ArcGIS, the Global Moran's Index of the housing price in Zhengzhou City and the calculation results are reported in Figure 12.

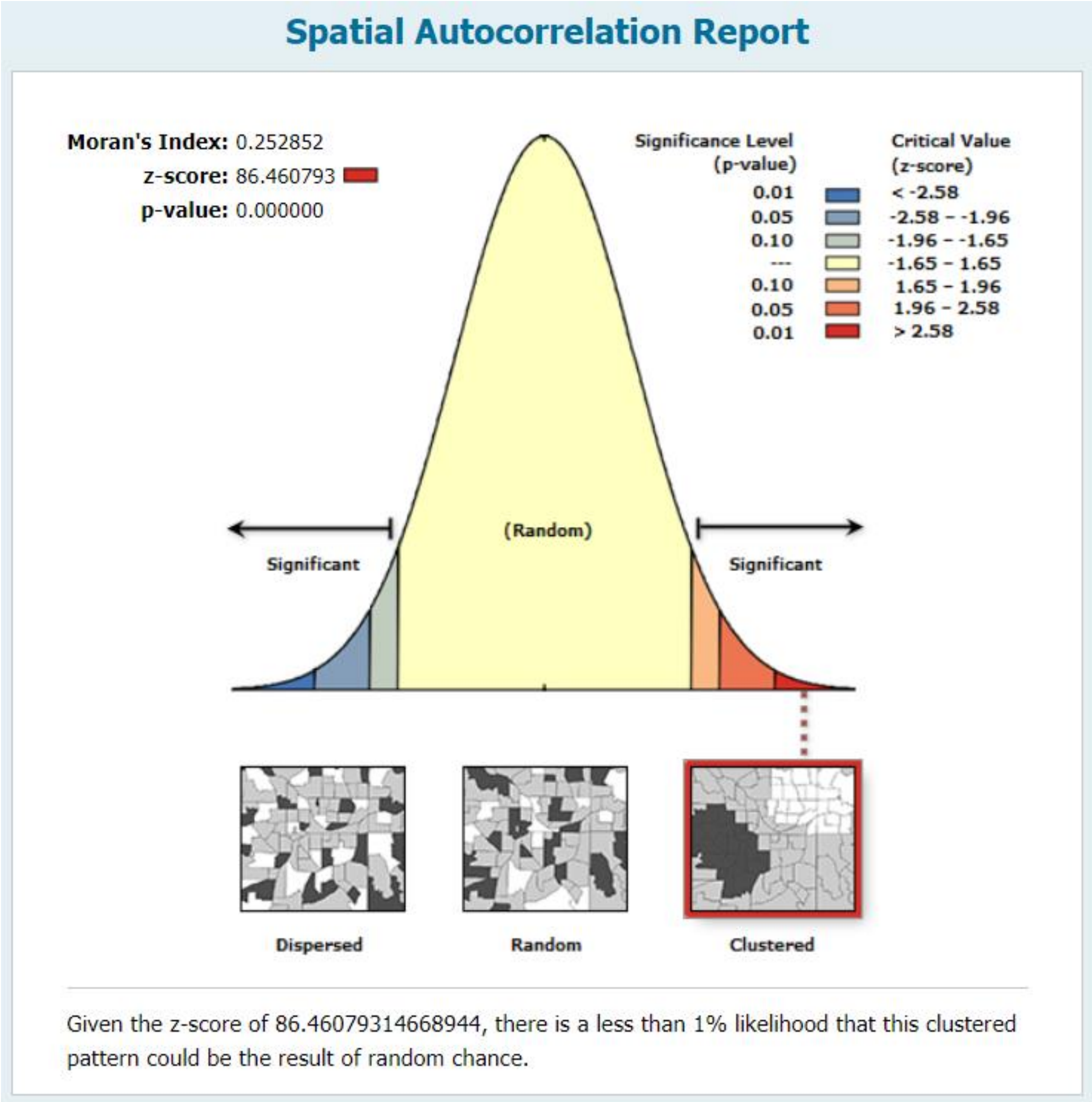

Figure 12 Spatial autocorrelation report of the transaction housing price

We find that the Moran's Index of housing price is 0.253, which is strongly positively correlated. With a z-score of 86.46, the housing prices show a significant spatial autocorrelation feature at a 0.000 significance level. The conclusion can be drawn as the overall distribution of housing prices in Zhengzhou City has a positive autocorrelation and shows a remarkable spatial high-high agglomeration and low-low agglomeration.

The global Moran's I index analysis shows that the spatial distribution of housing prices has spatial aggregation characteristics. Therefore, traditional regression analysis method (OLS) has limitations in urban housing price analysis and needs a massive paradigm shift to geographically weighted analysis. Also, the global spatial autocorrelation presupposes the homogeneous variation of the entire regional space and cannot effectively measure the spatial local heterogeneity. In regard to this, the local spatial autocorrelation analysis (Local Moran's Index) will be adopted to analyze the spatial heterogeneity of the housing price distribution in the study area.

4.1.3.2 Local spatial autocorrelation analysis (Local Moran's Index)

The local spatial autocorrelation analysis mainly analyses the distribution pattern of each attribute value in the heterogeneous region and measures the degree of local correlation between each position attribute value and the neighbouring area attribute value. Using the cluster and outlier analysis (Local Moran's Index) tool in ArcGIS, the local spatial correlation analysis of the housing price is carried out, and the significant level of 0.05 is selected to produce the corresponding LISA map (Figure 13).

The red points in the figure are the local high value (HH) agglomeration points, the light blue points are the local low value (LL) agglomeration points. The yellow points are the local high value (HL) outlier points, and the dark blue points are the local low value (LH) outlier points. The HL dots indicate that the houses are surrounded by those with lower prices, while the LH dots portray that they are neighbouring with houses with higher prices.

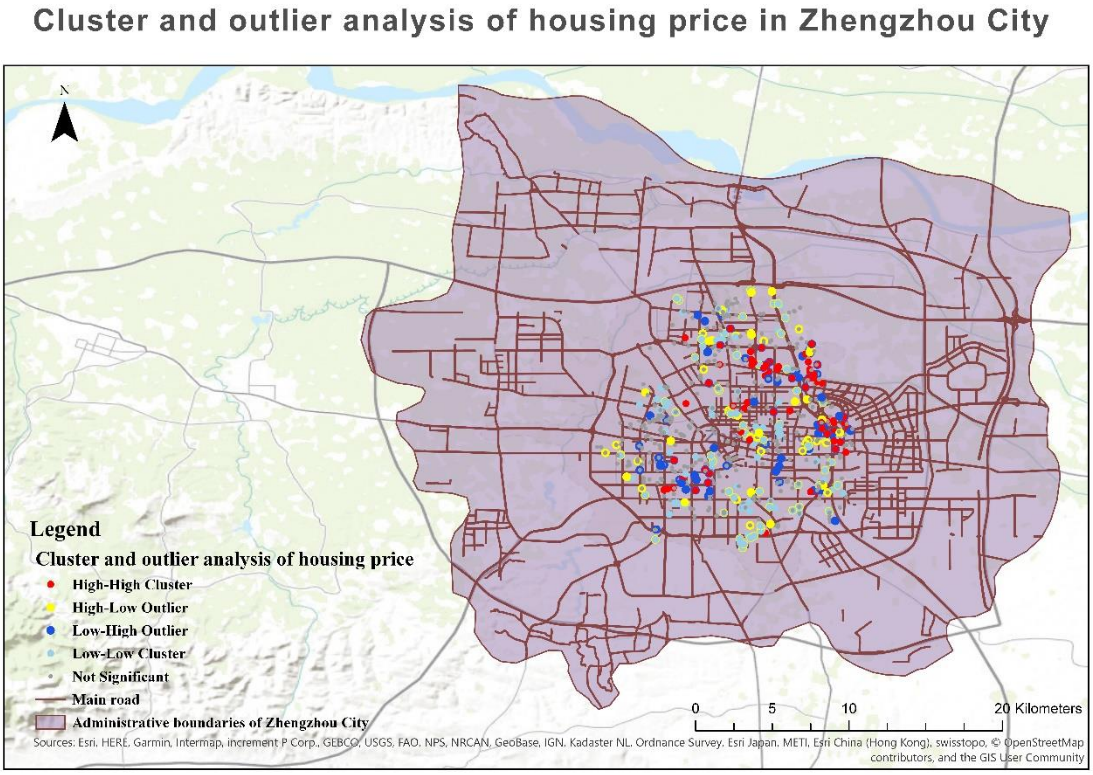

Figure 13 Cluster and outlier analysis of housing price in Zhengzhou City

In the LISA map of housing price in Zhengzhou City, there are 1963 red dots and 3013 light blue dots, occupying the 74.45% of the total sample points (Figure 14). The red dots mainly distributed in the east and northeast direction which is in line with the reality that housing prices in these districts have continued to grow in recent years. For a long time, the main development direction of Zhengzhou city is in these two directions. Also, with the development of Zhengdong New District, there are more and more house with higher prices being built there. The developers focus on building comfortable multi-storey (<=7) houses with ecological environment in the northeast of Zhengzhou City. While the light blue dots are concentrated in the southwest direction, which are located in the high-tech zone, economic development zone and airport area. They are mainly built for the working people and also there are fewer supporting policies for those areas to develop real estate industry. Those yellow points imply that they are built in recent years compared with the nearby houses so that they have higher prices than others. The dark blue points are those who have a longer house age than their neighbouring houses.

High-high points and Low-low points indicate that housing prices in the region are similar and agglomerate, while High-low dots and Low-high dots imply regional housing prices with partial spatial heterogeneity. Therefore, the regression analysis of housing prices in Zhengzhou City must take the characteristics of local spatial heterogeneity into account, and it is necessary to build a GWR estimation with spatial coordinate of each sample house.

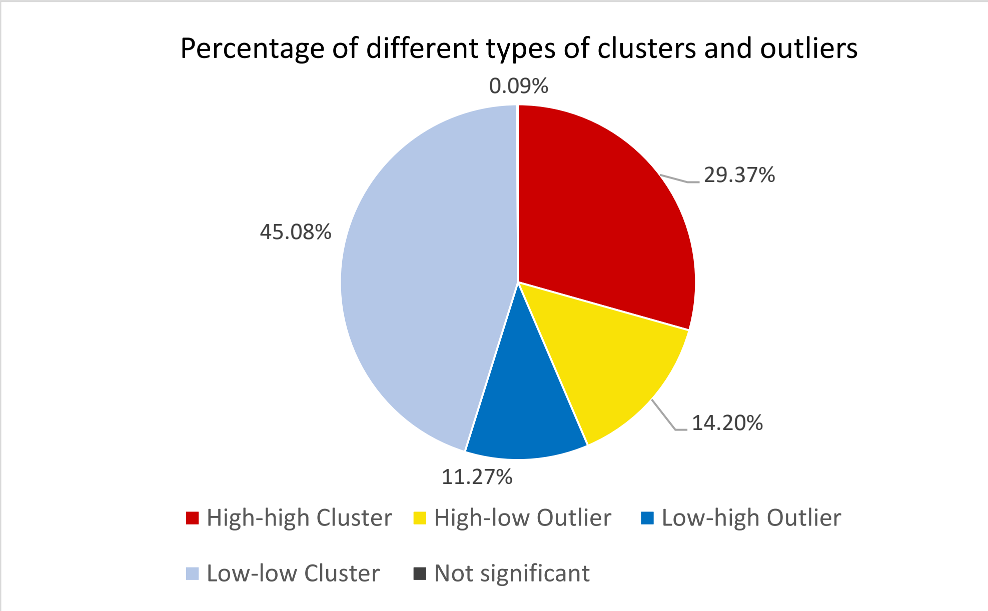

Figure 14 Percentage of different types of clusters and outliers

4.2 The analysis of second-hand housing attributes and housing price based on the Ordinary Least-Squares (OLS) estimation technique

An estimation is an expression of the system, process, thing, or concept being studied. While an econometrics estimation can be a mathematical equation or a quantitative or qualitative portrait of the interrelationship or causal correlation between variables in the system (Xie, 2013). Based on the results of inferential analysis, this part constructs a general linear regression (OLS) estimation to quantitatively interpret and analyze the influencing factors of housing prices.

4.2.1 Variables and descriptive statistics

The most general method to analyze the correlation between housing attributes and housing price is the Hedonic Price Model in the existing researches, which is to establish the relationship between variables by constructing general linear regression equations.

The variables selected in this paper have been introduced in Chapter 3. The main attributes and their quantifications are shown in Table 3. The expected impact refers to the correlation between each variable and the housing price. The "+" represents that the variable is positively correlated with the housing price, that is, the variable has a positive impact on the housing price; "-" represents the negative correlation, then the variable usually has a negative impact on the housing price. Among them, the dependent variable is the logarithmic functional form of housing price which is similar to a normal distribution.

Table 3 Definitions and expected impact of variables

Indicators	Definitions	Quantification	Expected impact
lgprice	Logarithmic functional form of housing transaction price	Number in RMB	N/A
lgsize	Logarithmic functional form of building size	Number in square meter	+
Floors	Which floor is the housing on	Number	+/-
Building type	Building type	Dummy variable: low-rise building=0 (<28 storeys), high-rise building=1(>=28 storeys)	+/-
Building orientation	If the house faces south or north	Dummy variable: South=1, North=0	+
Houseage	House age	2018 minus the building year	-
Plot ratio	Plot ratio of the neighborhood	Number	-
Green ratio	Green ratio of the neighborhood	Number	+
Dist_park	Distance to the nearest park	Kilometre	+
Dist_hospital	Distance to the nearest hospital	Kilometre	+/-
School district	If it is in a first-class primary school district	Dummy variable: Yes=1, No=0	+
Bus stops	Number of bus stops within 1 km	Number	+
Dist_subway	Distance to the nearest subway station	Kilometre	-
Dist_road	Distance to the nearest arterial road	Kilometre	-
Dist_Erqi	Distance to Erqi Square(CBD)	Kilometre	-
Dist_Jinshui	Distance to Jinshui River Park	Kilometre	-

+ means positive effect on housing price, – means negative effect on housing price

Table 4 shows the descriptive statistics for the dependent as well as the independent variables considered in the following OLS estimation. As what can be seen from the table, the overall sample size is 8864. As for the housing price, the minimum price is 378,449 and the maximum price is 2,000,000, which indicates a relatively big range of the housing price in Zhengzhou City. After being transformed into a logarithmic form, the standard deviation is shrunk to 0.27. All the sample points are located from 113.600 to 113.800 in the longitudinal and from 34.690 to 34.840 in latitude, which are in line with the scope of Zhengzhou City.

In terms of the structural characteristics, the building size ranges from 60m^2 to 144 m^2, which is in line with the construction standard of commercial residential house in China. And the logarithmic form of building size varies from a minimum of 4.095 to a maximum of 4.969. Half (54.31%) of the sample houses are low-rise buildings, and 3054 houses are high-rise buildings with more than 28 stories. The mean of the building orientation is 0.19, which indicates that most of the houses are facing to the south and have full sunlight. Besides, the houses selected in the sample were built between 1992 and 2016, so that the minimum of the house age is 3 and the maximum of the house age is 27.

In respect of the neighborhood characteristics, the average numbers of plot ratio and green ratio are 3.26 and 0.35 respectively, which are similar to the results of the research of Zhengzhou City of Dong(2017). For the distance to the nearest park and hospital, namely, the average distances are both 0.97 kilometers. When it comes to the indicator of school district, the mean is 0.08, which implies there are very few (8.12%) houses located in a first-class primary school district. This may provide a hypothesis that the indicator “first-class primary school district” will have an important impact on the housing price.

As for the locational characteristics, the indicators Bus stops, distance to subway station, and distance to the main road are used to measure the convenience of transport. With a mean value of 1.91, most sample houses (88.27%) are furnished with at least one bus stop within 1 km. The average number of the distance to nearby subway station is 2.37, with a higher standard deviation of 1.70, compared with other indicators. Since there are only two subway lines in Zhengzhou City, some houses in this study are located far away from metro stations. While the distance to the main road seems better because the average number is only 0.19 km, which indicates a better convenience of transportation. For the distance to the CBD, the minimum is 0.087 km and the maximum is 9.805 km (located in the outskirt of Zhengzhou City) with a mean of 5.49. This is because Zhengzhou City is designed as a monocentric city. Severed as a large-scale urban landscape, the average distance to Jinshui River Park is 5.47 km with a highest distance of 9.109 km and a lowest number of 0.313 km.

Table 4 Descriptive statistics of variables

Variable	Observation	Mean	Standard deviation	Minimum	Maximum
lgprice	6,684	13.76	0.27	12.840	14.510
lgsize	6,684	4.57	0.22	4.095	4.969
Floors	6,684	13.72	8.43	1.000	34.000
Buildingtype	6,684	0.46	0.50	0.000	1.000
Buildingorientation	6,684	0.96	0.19	0.000	1.000
Houseage	6,684	7.71	2.85	3.000	27.000
Plotratio	6,684	3.26	1.19	0.150	10.000
Greenratio	6,684	0.35	0.07	0.100	0.580
DIST_park	6,684	0.97	0.78	0.000	3.900
DIST_hospital	6,684	0.97	0.53	0.011	2.598
School district	6,684	0.08	0.27	0.000	1.000
Bus stops	6,684	1.91	1.20	0.000	5.000
DIST_subway	6,684	2.37	1.70	0.088	9.600
DIST_mainroad	6,684	0.19	0.20	0.000	1.196
DIST_CBD	6,684	5.49	1.80	0.087	9.805
DIST_Jinshui park	6,684	5.47	1.68	0.313	9.109
LONG	6,684	113.70	0.04	113.600	113.700
LAT	6,684	34.77	0.04	34.690	34.840

4.2.2 Correlation analysis

In this part, 16 variables were thrown into correlation analysis, and the results generated by STATA 14.2 are shown in Table 3. The value of correlation coefficient ranges from -1.000 to +1.000 and the sign implies the direction of the relationship.

It is noted that the floors have a moderate positive relationship with the building type because the higher value in building type refers to a higher building. Also, the floors and the house age are negatively correlated (-0.226). This can be explained like, with the growing up trend of population, more and more residential houses in recent years tend to be build higher. Besides, the houses located on a higher floor implies there is a higher plot ratio of the corresponding neighborhood (0.250). Anecdotally, the indicators "floors" and "building type" tend to have negative correlation with all the other indictors measuring distance. This provides an indication that higher buildings share a better convenience with the nearby public services. Compared with the strength of other indicators, the house age has a quite strong correlation with the indicator of distance to CBD (-0.171) and the large-scale landscape Jinshui River Park (-0.206). Since the older houses were built around the central part of Zhengzhou City, where locates the CBD and the Jinshui River Park. We also find the strong positive relationship between the distance to subway station and distance to CBD (0.434) and Jinshui River Park (0.375), because the two subway lines are set around the main district of Zhengzhou City.

The correlation coefficient between the indicator distance to CBD and the indicator distance to Jinshui River Park reaches 0.903 because the actual distance between these two coordinates is only 1.3 km. In order to avoid the inaccuracy brought by the multicollinearity, the indicator of Jinshui River Park will be thrown into the regression estimation later after testing the robustness of the estimation and variance inflation factors.

Table 5 Correlation analysis of variables

	lgprice	lgsize	Floors	Buildingtype	Orientation	House age	Plot ratio	Green ratio	DIST park	DIST hospital	School district	Bus stops	DIST subway	DIST road	DIST CBD	DIST Jinshui park
lgprice	1.000															
lgsize	0.739*	1.000														
Floors	0.059*	-0.043*	1.000													
Buildingtype	0.141*	-0.027*	0.415*	1.000												
Orientation	0.236*	0.232*	-0.001	0.010	1.000											
Houseage	-0.089*	0.039*	-0.226*	-0.372*	-0.080*	1.000										
Plotratio	0.069*	-0.014	0.250*	0.420*	0.007	-0.242*	1.000									
Greenratio	0.115*	-0.043*	-0.025*	0.019	0.046*	-0.003	-0.154*	1.000								
DIST park	-0.029*	-0.014	-0.048*	-0.099*	0.006	-0.111*	-0.106*	0.007	1.000							
DIST hospital	-0.024*	0.029*	-0.058*	-0.093*	-0.029*	0.028*	-0.059*	-0.019	0.237*	1.000						
School district	0.022	-0.051*	0.072*	0.180*	0.017	-0.083*	0.152*	0.155*	-0.024	-0.236*	1.000					
Bus stops	0.029*	-0.016	0.050*	0.139*	-0.035*	0.008	0.064*	0.029*	-0.104*	-0.157*	0.156*	1.000				
DIST subway	-0.079*	0.019	-0.025*	-0.094*	-0.008	-0.122*	-0.109*	-0.055*	0.354*	0.090*	-0.013	-0.048*	1.000			
DIST road	-0.011	0.013	-0.078*	-0.120*	-0.036*	0.064*	-0.131*	0.113*	0.008	0.298*	-0.088*	-0.062*	-0.040*	1.000		
DIST CBD	-0.042*	0.022	-0.100*	-0.162*	0.022	-0.171*	-0.274*	0.131*	0.542*	0.369*	-0.038*	-0.031*	0.434*	0.282*	1.000	
DIST Jinshui park	-0.078*	0.043*	-0.082*	-0.142*	0.015	-0.206*	-0.234*	0.041*	0.445*	0.396*	0.018	0.003	0.375*	0.228*	0.903*	1.000

Notes: * means the 0.05 significance level.

4.2.3 Ordinary Least-Squares (OLS) regression

Table 6 reports the results of the OLS estimation including the structural characteristics, neighborhood characteristics and locational characteristics. The overall explanatory power of this model is quite strong, measured by the R^2 (0.6265), indicating that 63% of the housing price in Zhengzhou City can be explained by the independent variables in this model. In the test of each regression coefficient, if the P-value is less than 0.01, which implies that the variable is significant and has a stronger interpretation of the housing price. In this model, 13 out of 15 variables are significant at a at least 0.05 significance level and most of them are significant at a 0.0000 significance level.

Table 6 Ordinary Least-Squares (OLS) regression results (dependent variable: lgprice)

Variable	Coefficient	t-value	VIF
Structural characteristics			
lgsize	0.901***	96.90	1.07
Floors	0.001***	3.33	1.23
Building type	0.044***	8.62	1.58
Building orientation	0.064***	6.01	1.08
House age	-0.010***	-10.35	1.37
Neighborhood characteristics			
Plot ratio	0.001	0.51	1.39
Green ratio	0.457***	14.59	1.16
DIST park	0.007**	2.24	1.56
DIST hospital	0.015***	3.38	1.42
School district	0.029***	3.67	1.20
Locational characteristics			
Number of bus stops	0.007***	4.32	1.08
DIST subway station	-0.012***	-8.87	1.33
DIST main road	-0.018	-1.46	1.25
DIST CBD	0.034***	11.20	7.59
DIST Jinshui River Park	-0.051***	-16.73	6.39
Constant	9.546***	213.20	
Sample size	**6684**		**Mean VIF**
$\mathbf{R^2}$	**0.6265**		
F(17, 6666)	**728.55**		**2.05**
Prob > F	**0.0000**		

Notes: * significane at the 0.000 level, ** significane at the 0.05 level.**

After constructing the OLS regression, several relative tests needed to be done. Multicollinearity is often considered in this stage and the Variance Inflation Factor (VIF) is a common method to detect if there is multicollinearity in a model. As reported in Table 6, the overall VIF is 2.05 and the VIF value of all the independent variables are less than 10. This indicates that there is no multicollinearity concern in this model. Also, an OLS estimation asks that the variance in the residuals must be homoscedastic. This is also tested in STATA and the chi^2 is 9.22 with a P-value of 0.0024, which implies that the residuals are skewed. And a robust order is called to adjust this heteroskedasticity. A normality test is needed to judge if the residuals of the model distributed normally. Shown in Figure 15, this model also passed the tests of normality and doesn't have omitted variables. In a nutshell, the results of this OLS estimation are reliable.

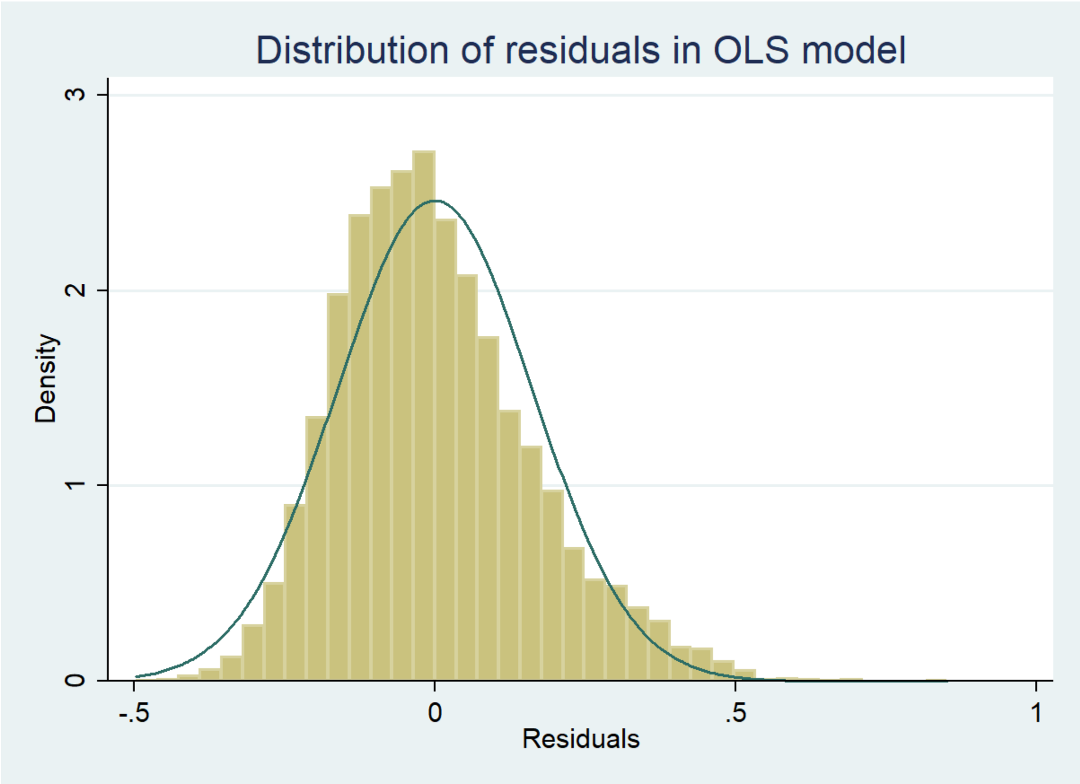

Figure 15 Kernel density estimate of residuals in OLS estimation

According to the calculation results, the linear regression model of Zhengzhou residential price can be expressed as follows:

$$\begin{aligned}\mathrm{Ln}P = 9.546 &+ 0.901\mathrm{Ln}X_1 + 0.001X_2 + 0.044X_3 + 0.064X_4 - 0.010X_5 + 0.001X_6 \\ &+ 0.457X_7 + 0.007X_8 + 0.015X_9 + 0.029X_{10} + 0.007X_{11} - 0.012X_{12} \\ &- 0.018X_{13} + 0.034X_{14} - 0.051X_{15}\end{aligned}$$

As can be seen from the formula, the relationship between the housing attributes and the housing price could be interpreted as follows:

(1) Structural characteristics. The building size has a strongest positive impact on the housing price with its coefficient of 0.901. To be specific, every 1% improvement of building size will increase the housing price by 0.9%. Also, the indicator floors and building orientation have positive influence on housing price, which are in line with the expected impact. Here, the building type is a binary variable. Compared with the low-rise buildings, high-rise buildings (>=28 storeys) tend to have higher prices. Besides, with a coefficient of -0.010, a house with longer house age will reduce its housing price.

(2) Neighborhood characteristics. Analysis results show that, all the neighborhood characteristics have positive influence on housing price. if green ratio increases by 0.1, price is expected to increase by 4.57%. While the indicators of distance to park and hospital have positive coefficients, which means that the houses that are far away from these two public services have higher prices. This may because the noise brought by the public services and also reveal the suburbanization phenomenon in Zhengzhou City to some degree. The indicator school district has a positive coefficient of 0.029 and the corresponding P-value is 0.000, which indicates that the school district has a significant impact on housing price. It is apparent that people have a clear preference for the surrounding education package when buying a residence.
(3) Locational characteristics. The package of indicators measuring the traffic convenience show the same expected impact as we wish. The bus stops have a positive influence on housing price and this indicator is significant at a 0.000 significance level. With a negative coefficient of -0.012, people like to pay a higher price for those houses located near a subway station in case to travel around quickly and easily. Just like the indicator distance to hospital, the distance to CBD also shows a positive impact on housing price. To be exact, when the distance increases 1km, the housing price is expected to go up by 3.4%. The large-scale landscape Jinshui River Park influences the housing price positively. Jinshui River Park has fresh air and pleasant scenery, which will inevitably have a significant premium effect on the surrounding housing prices.

4.2.4 Spatial autocorrelation of residuals in OLS estimation

The OLS estimation only provides a certain degree of explanation in the global average sense, and 37% of the housing price changes have not been explained. This is partly because the global model assumes that the continuous change in housing prices in space is stable, but actually its changes are non-stationary and needed to put tested in a GWR estimation (Brunsdon, et al., 2002). From the residual spatial distribution analysis of the OLS estimation, the following report shows that there is significant spatial autocorrelation in the residuals (see Figure 15). The z-score of the Moran's Index reaches 68.76 and it is significant at a 0.0000 significance level, indicating that the regression results of the OLS estimation have large deviations, as a global model. To avoid the misleading results caused by the above OLS estimation, a GWR estimation will be deployed in the next part.

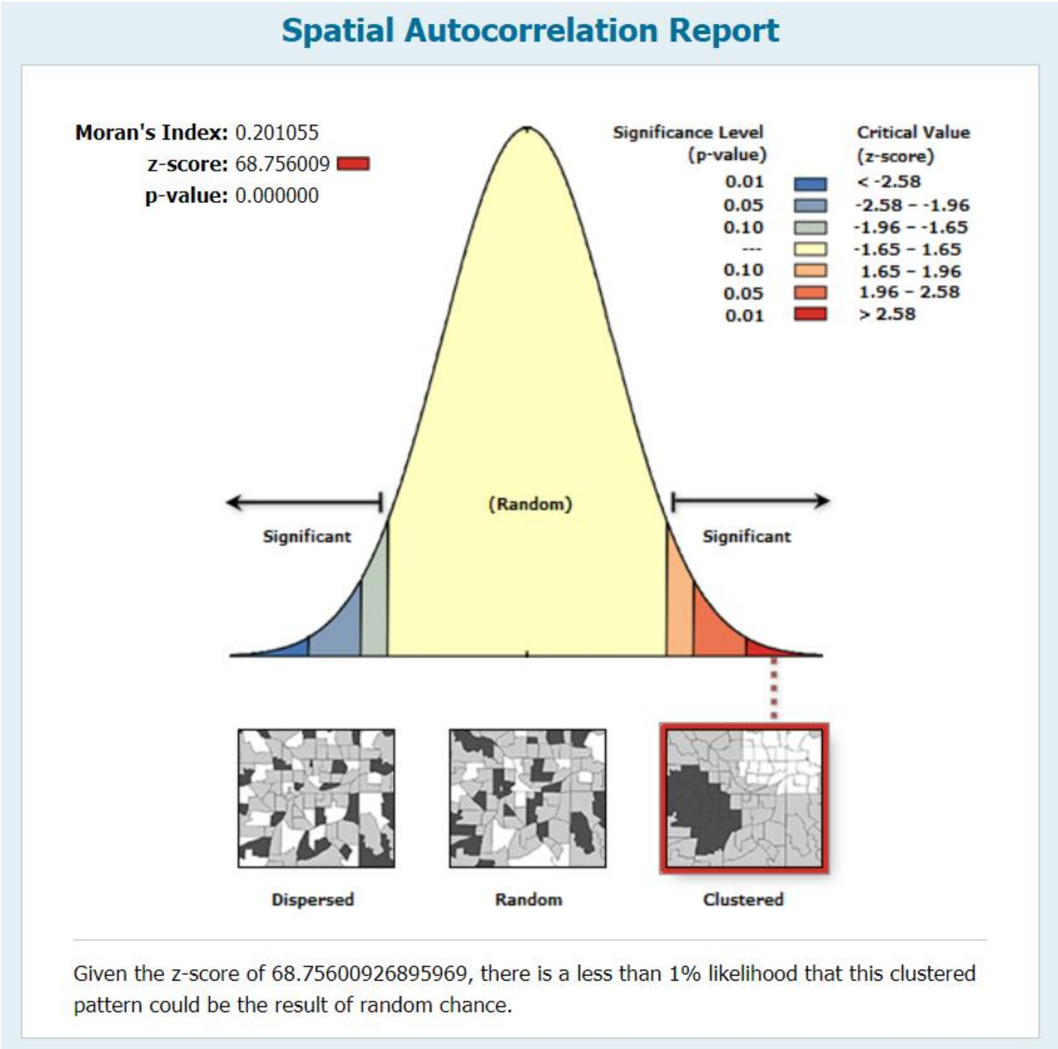

Figure 16 Spatial autocorrelation report of the residuals of OLS estimation

4.3 The analysis of spatial heterogeneity in the price composition of the second-hand housing based on Geographically Weighted Regression (GWR) estimation technique

From a geospatial perspective, location is a key factor affecting housing prices due to the spatial heterogeneity of housing prices. Because the housing price distribution is spatially heterogeneous and statistically non-stationary, the Geographically Weighted Regression estimation is usually used for the analysis between housing attributes and housing prices (Brunsdon, et al., 2002).

4.3.1 Model construction and inspection

In the GWR estimation, the selected dependent variable and independent variables are the same as those in Section 4.2. The price of the ith house is P_i, and the corresponding coordinate is (u_i, v_i). X_{ij} represents each housing attribute and the β_i refers to the coefficient of each characteristic. In a nutshell, the formula could be generated as follows:

$$
\begin{aligned}
P_i = \beta_0(u_i,v_i) &+ \beta_1(u_i,v_i)X_{i1} + \beta_2(u_i,v_i)X_{i2} + \beta_3(u_i,v_i)X_{i3} + \beta_4(u_i,v_i)X_{i4} \\
&+ \beta_5(u_i,v_i)X_{i5} + \beta_6(u_i,v_i)X_{i6} + \beta_{27}(u_i,v_i)X_{i7} + \beta_8(u_i,v_i)X_{i8} \\
&+ \beta_0(u_i,v_i)X_{i0} + \beta_{10}(u_i,v_i)X_{i10} + \beta_{211}(u_i,v_i)X_{i11} + \beta_{12}(u_i,v_i)X_{i12} \\
&+ \beta_{13}(u_i,v_i)X_{i13} + \beta_{14}(u_i,v_i)X_{i14} + \beta_{15}(u_i,v_i)X_{i15}
\end{aligned}
$$

Also, as what has been mentioned in the Section 3.6.3, the Akaike Information Criterion (AIC) will be used to test the goodness-of-fit of a GWR estimation in the following part.

4.3.2 Geographically Weighted Regression (GWR) estimation method

In this paper, the data of housing prices in Zhengzhou City will be entered into the GWR 4.0 software. Since the sample points in this paper are not distributed discretely, the Fixed Gaussian function is selected as the spatial kernel type, and the AICc method is used to determine the final bandwidth. Table 7 is the descriptive statistics of regression coefficients of each housing attribute. The GWR estimation explains 68.4% of the changes in housing price, with a set of local R^2 for each observation. The local R^2 ranges from 0.450 to 0.822, and 50% of them are higher than 0.7, implying a good goodness-of-fit of this GWR estimation. Besides, the mean value of the Cook's Distance is 0.000, reporting that there are no outliers in this estimation. In terms of the dependent variables, the regression coefficients of each explanatory variable can be positive or negative, indicating different housing attributes will have both value-added and value-decreased effects on housing price. Since various urban development factors of different locations are not balanced, the influence of these factors on housing prices has significant spatial heterogeneity. Besides, only significant (at least at a 0.1 significance level) indicators will be specifically analysed in the following part.

Table 7 Descriptive statistics of GWR estimation coefficients

Variable	β mean	β minimum	β maximum	β Range	β standard deviation
lgsize	0.885^{***}	0.789	0.998	0.209	0.045
Floors	0.001	-0.002	0.003	0.005	0.001
Building type	0.019	-0.077	0.080	0.157	0.026
Building orientation	0.057^{*}	-0.040	0.135	0.175	0.038
House age	-0.012^{***}	-0.024	0.004	0.027	0.006
Plot ratio	-0.006^{*}	-0.045	0.029	0.074	0.017
Green ratio	0.366^{***}	-0.328	0.811	1.139	0.229
DIST_park	0.028^{*}	-0.178	0.175	0.354	0.061
DIST_hospital	-0.014^{*}	-0.159	0.197	0.356	0.058
School district	-0.007	-0.255	0.180	0.435	0.080
Bus stops	0.005	-0.024	0.044	0.068	0.013
DIST_subway	-0.012^{**}	-0.048	0.028	0.076	0.013
DIST_mainroad	0.021^{*}	-0.418	0.263	0.681	0.137
DIST_CBD	0.132^{*}	-1.974	3.618	5.592	0.508
DIST_Jinshui park	-0.149^{*}	-3.648	2.071	5.718	0.513
Intercept	9.707^{***}	6.194	11.536	5.341	0.573
Local R-squared	0.679	0.450	0.822	0.372	0.073
Cook's Distance	0.000	0.000	0.045	0.045	0.001
Adjusted R-squared	**0.684**				

AICc	-6262.754

β : * significane at the 0.000 level, ** significane at the 0.05 level, * significane at the 0.1 level.**

(1) Local R^2

The range of local R^2 ranges from 0.45 to 0.82, and it indicates that 79% of the sample points has a higher goodness-of-fit than the same points in OLS estimation (Figure 17). Based on the inverse distance weighted (IDW) technique, Figure 18 shows an overall distribution of the local R^2 of GWR estimation in Zhengzhou City. The geographical map delineates that, except the points located in a small area of northwest city, other sample dots have relative higher values of R^2. The lower R^2 in the northwest part may because the urban space structure is based on the concentric circle model when constructing the Hedonic Price Model, only one core of the city center is selected as the basis for calculating the characteristic variables. The region with low local R^2 may be a potential subcentre of the city, which reveals to some extent the urban spatial structure of Zhengzhou City has a polycentric model development trend. In reality, the North Longhu area in the northwest locates many Grade-A neighbourhoods. The building quality, appearance style, residential technology and living comfort have all achieved top level in North Longhu area. Equipped with higher green ratio, limited plot ratio, education package, medical care, this area has a higher overall housing price than other areas. This is in line with the implied trend shown in the geo-map.

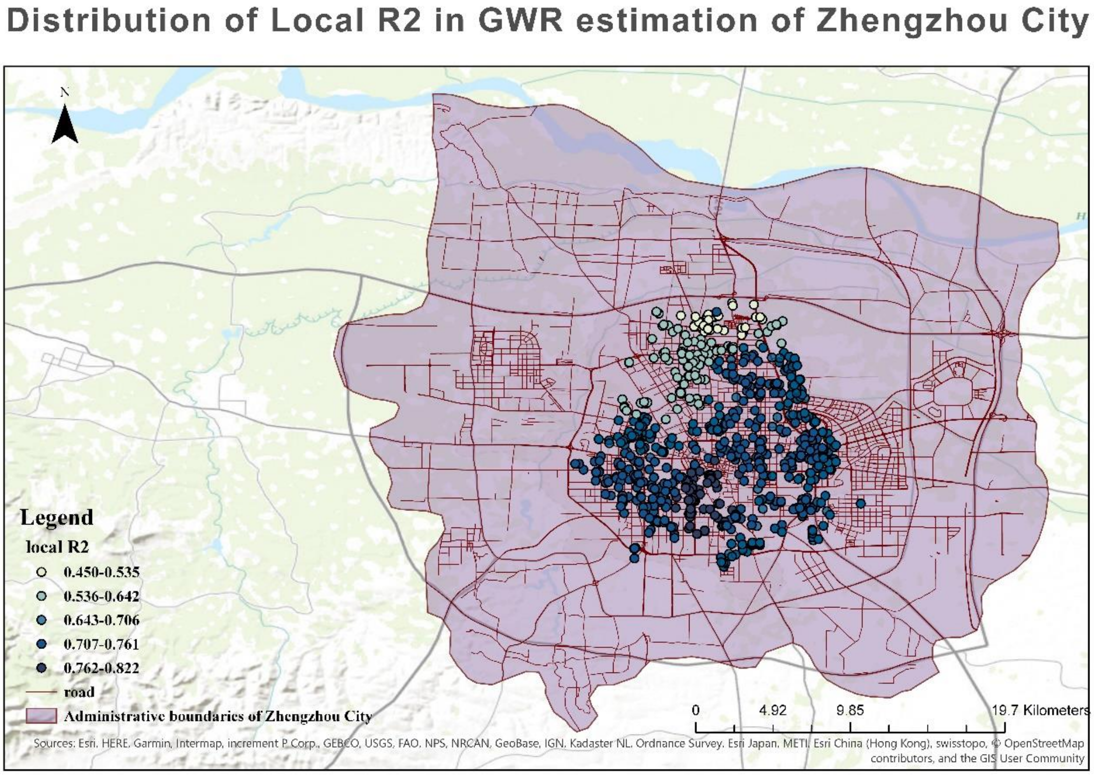

Figure 17 Distribution of Local R^2 in GWR model

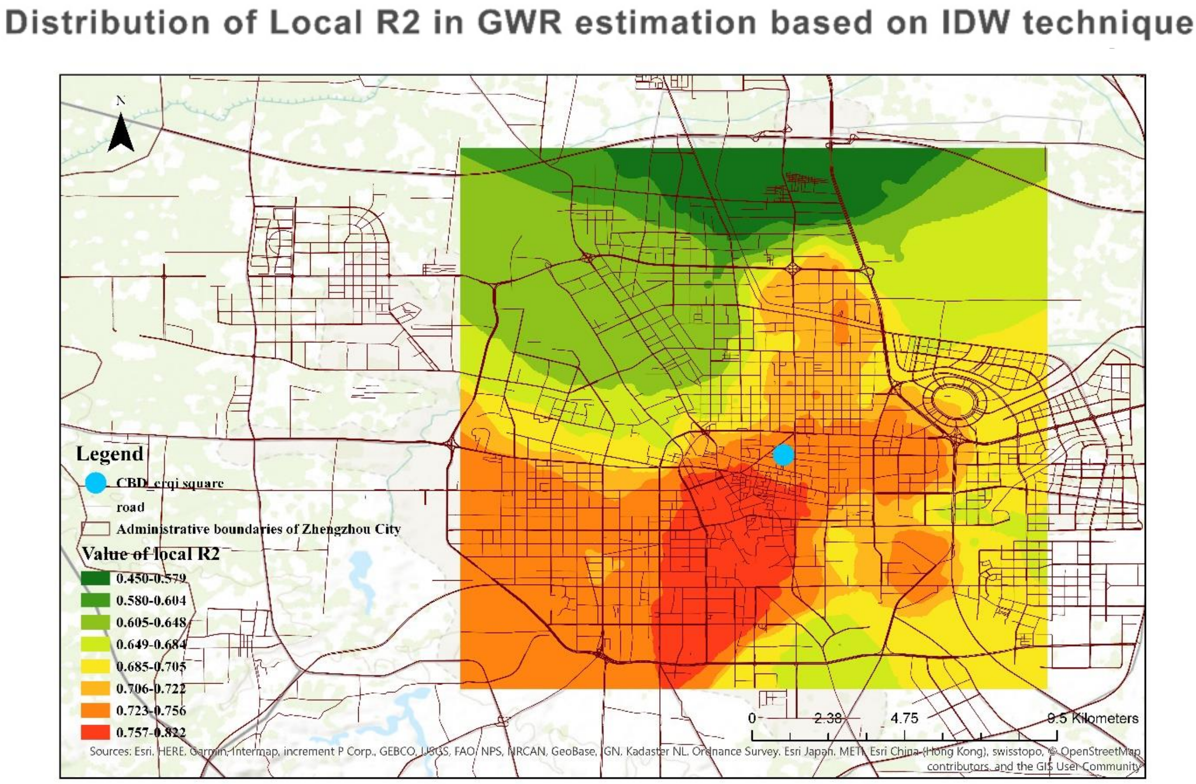

Figure 18 IDW distribution of Local R^2 in GWR estimation of Zhengzhou City

(2) Building size.

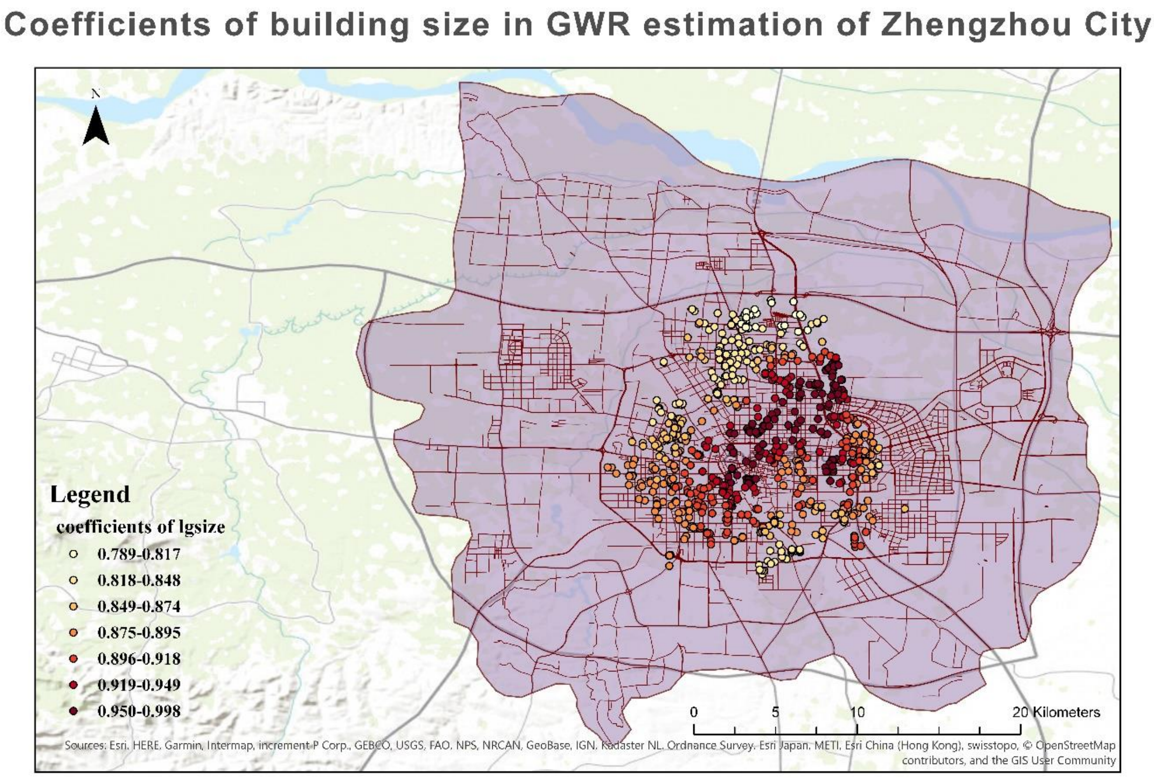

Figure 19 Distribution of coefficients of building size in GWR estimation

Combining the results from Table 7 and Figure 19, local coefficient of the logarithmic functional form of building size ranges from 0.789 to 0.998. All the coefficients are positive which indicate that the bigger the house, the higher its price will be. The highest positive elasticity between the building size and housing price can be observed in the central city and the eastern part of the east. Since the land prices in these areas are higher, the houses here have more elasticity to size than other areas. The smallest coefficients concentrated in the outside areas, revealing that houses far away from the city center are not so sensitive as those located in the central area.

(3) Building orientation

As can be seen from Table 7 and Figure 20, the coefficients of building orientation increase from a negative value of -0.040 to a positive value of 0.035. Distributed oppositely to building size, the small values concentrated in the central area, while the highest numbers located in the southern and east northern part of the city. Since most of the houses in the main urban areas are old houses with well-equipped infrastructure and other kinds of resources, the impact of housing orientation on residential prices is relatively small compared with other indicators. However, there are more new-build neighbourhoods in the surrounding areas, and residents have higher requirements for housing orientation.

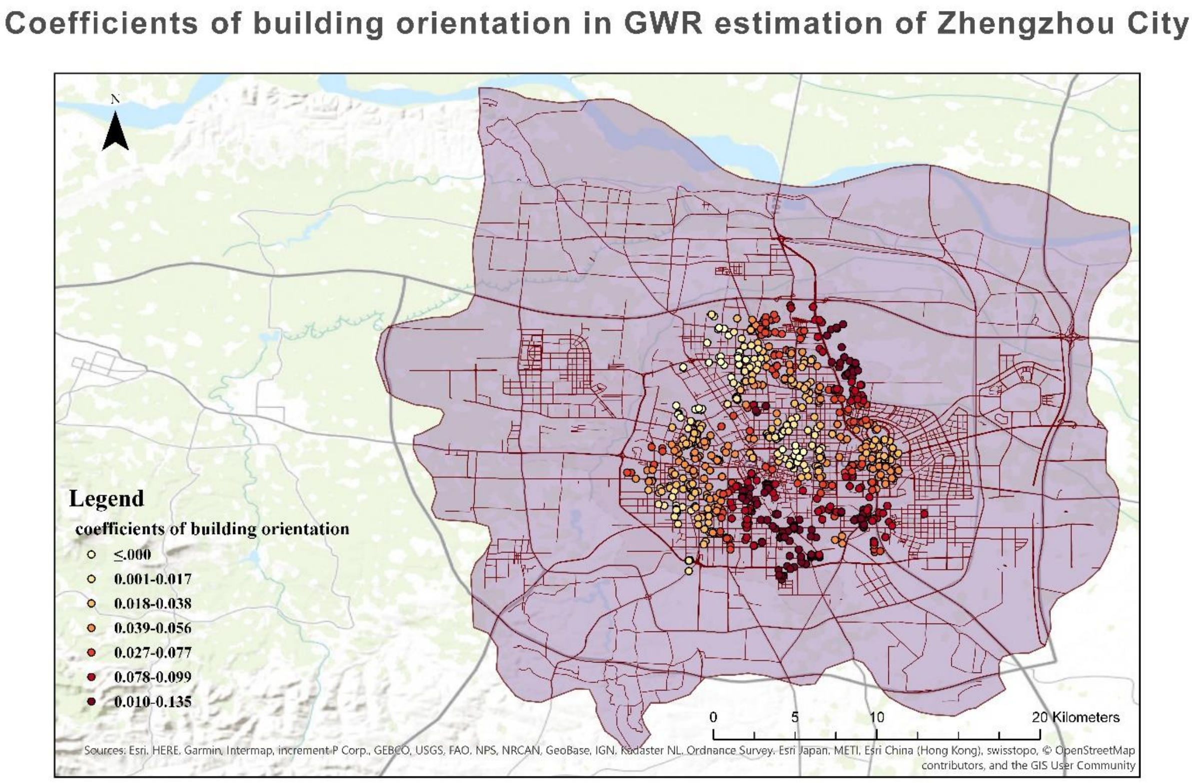

Figure 20 Distribution of coefficients of building orientation in GWR estimation

(4) House age.

Ranged from -0.024 to 0.004 (Table 7), more than 98% of the local coefficients of house age indicate its negative effect on housing price, which is in line with expectation. Based on Figure 21, we can find that the middle-value coefficients located in the central areas. The majority of the housing in the downtown area is more than 10 years old. However, due to the good location and the surrounding infrastructures, the impact of the house age on the residential price is reduced. The absolute value of the coefficients in the east, southeast and southwest are larger.

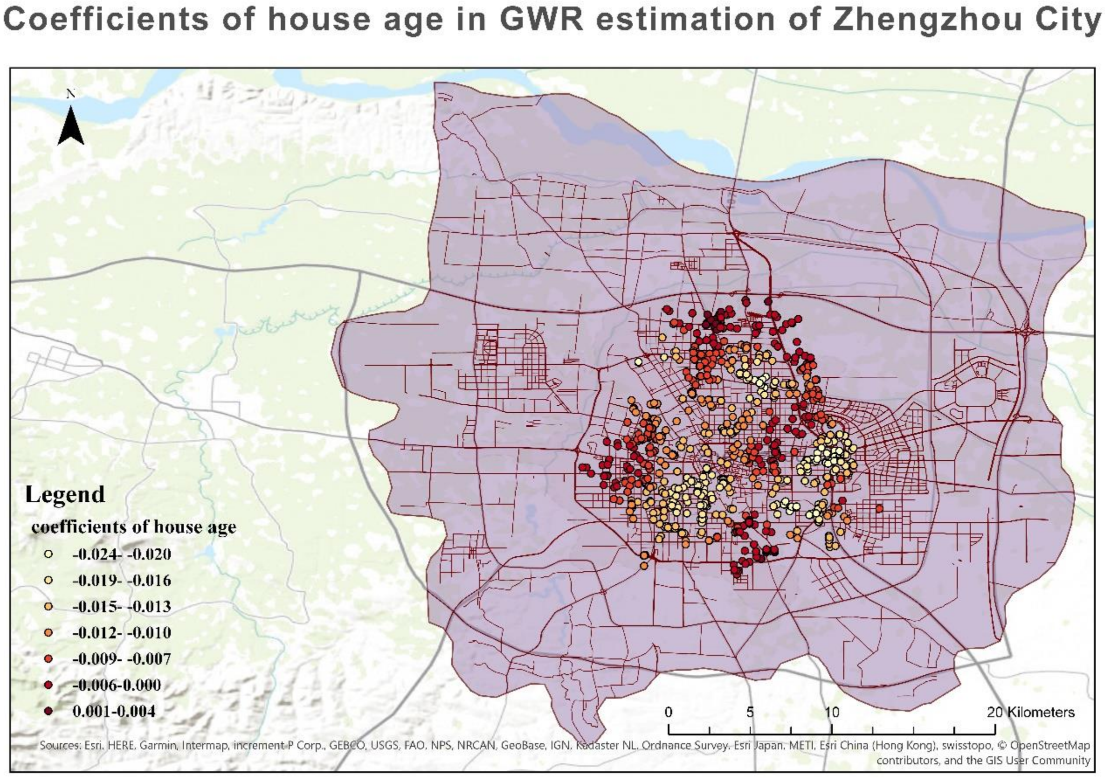

Figure 21 Distribution of coefficients of house age in GWR estimation

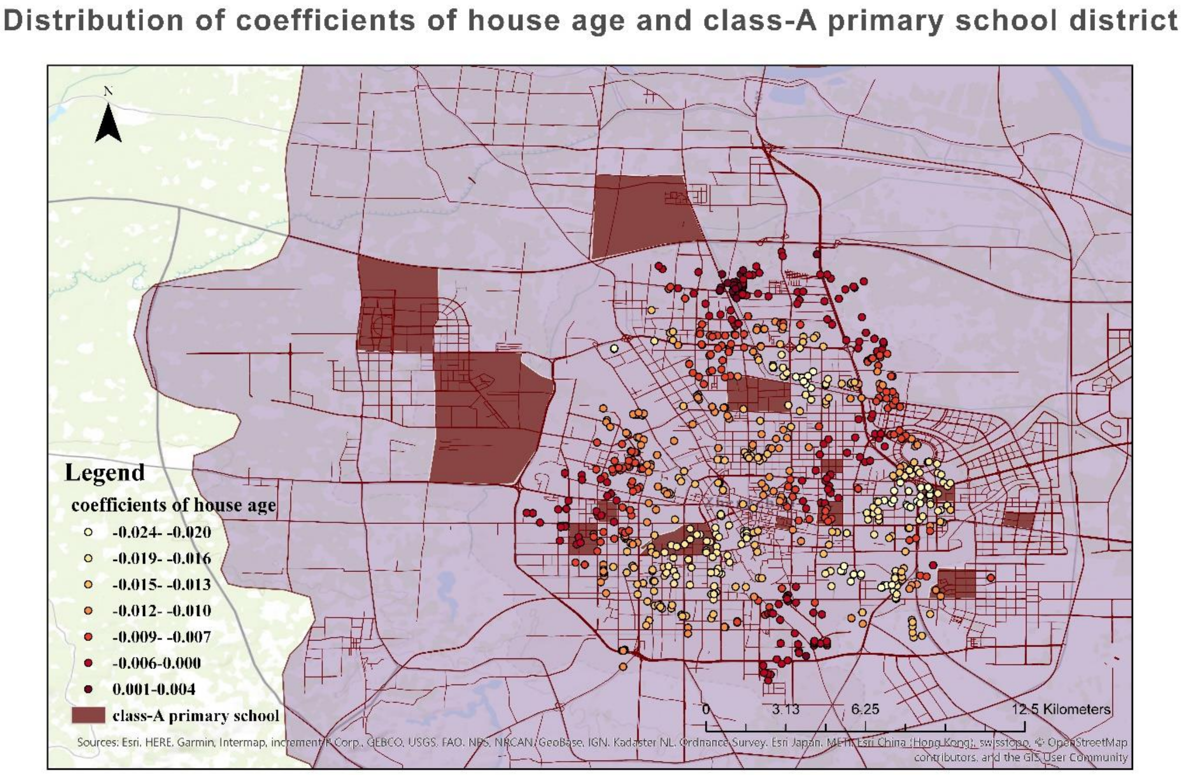

Figure 22 Distribution of coefficients of house age and class-A primary school district

Surrounded by a good educational package (class-A school district, see Figure 22), these areas attract more developers to invest in the construction of real estate. Compared with old houses, people prefer new real estate, which leads to higher impact of house age on housing price in this area. In the peripheral areas of Zhengzhou City, the house age seems to have no significant effect on housing price since the coefficient is almost equal to zero. Since this area is far from the main city, and the surrounding facilities are not complete, compared with house age, buyers focus on other housing attributes when buying a house. In a nutshell, house age is not a sensitive indicator in these areas.

(5) Plot ratio.

The coefficient of plot ratio ranges from -0.045 to 0.029 and it is significant at a 0.1 significance level. According to its statistical results, more than 65% of the coefficients are negative, so we can say that plot ratio generally has a negative impact on housing price. From the perspective of reality, a higher plot ratio will reduce the comfortable sensation of residents as it indicates there are more buildings or higher buildings within the neighborhood. However, if the plot ratio is too low, it implies an unreasonable plan of land use. From the results of Figure 23, in the direction from northwest to southeast, the coefficients of plot ratio first increase from a higher negative value (-0.046) to zero and then increase to a higher positive value (0.029). Since some high-grade neighbourhoods concentrated in the west northern part of the city and people prefer a more incompact residence, the residential price is sensitive to the plot ratio. While in the central city, because the relevant departments control the plot ratio of buildings, there is not too much variation of plot ratio in this area, the influence of this indicator is around zero. Look at the east southern part, many high-rise buildings or super high-rise buildings were built in recent years, so they tend to have higher prices. Therefore, the plot ratio seems to have a positive impact on housing price in these areas.

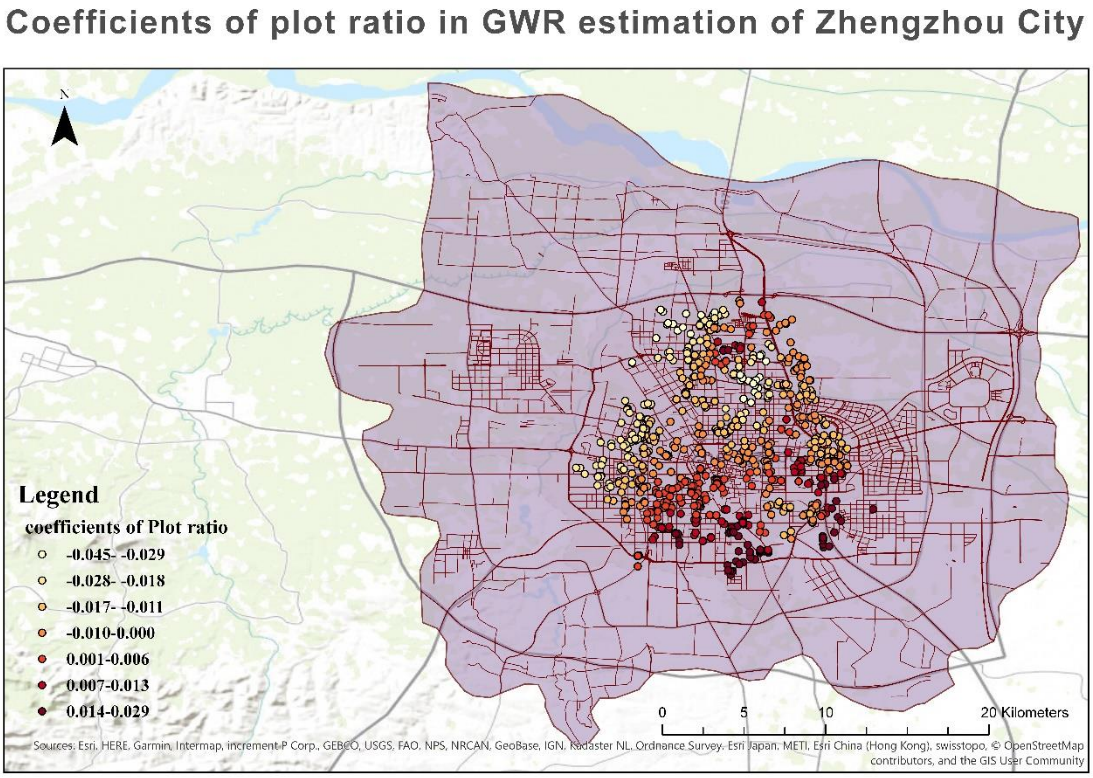

Figure 23 Distribution of coefficients of plot ratio in GWR estimation

(6) Green ratio

The coefficient of green ratio, ranging from -0.328 to 0.811, is significant at a 0.01 significance level. The majority (91%) coefficients of the sample points are positive which indicates the green ratio has a positive impact on housing price (Figure 24). Among them, the green ratio has a negative impact on the eastern and southwestern Zhengzhou City, denoting that the residential price decreases with the increase of the green ratio. Since the construction land in this area is relatively small and the land price is high, the increase in the green ratio will reduce the available building area, which leads to an increase in construction costs and changes the normal correlation between the housing price and the green ratio.

The stronger positive impacts locate in the northwest and southeast areas in Zhengzhou City. Due to the fact there are high-grade residential buildings in the northwest urban area, the residents are pursuing a comfort living environment, they prefer a greening community. While the southeast region is far from the urban area, with a relative late development time, and the surrounding infrastructure is not perfect. Therefore, the green ratio, one of the important factors of the quality of the residential environment, has become an important factor for buyers when choosing a house.

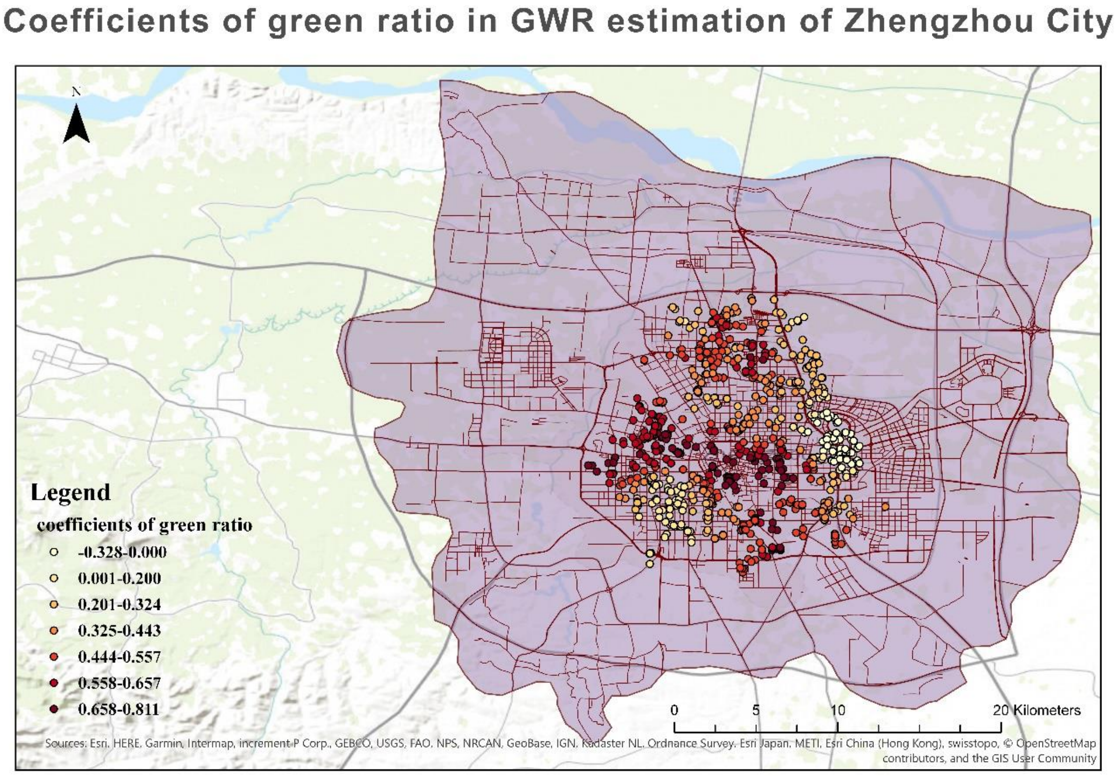

Figure 24 Distribution of coefficients of green ratio in GWR estimation

(7) Distance to park.

The distance to park, has a minimum coefficient of -0.178 and a maximum coefficient of 0.175. The areas with a negative coefficient are concentrated in the west northern neighbourhoods in Zhengzhou City (Figure 25), indicating that locating near a park will increase the housing price. As what has been mentioned above, there are more high-grade buildings here and people prefer to a better living environment. Parks could provide fresh air and relax people after a day of working. As a result, living near a park increases the housing price in these areas. However, the central city and the Zhengdong New District, with negative coefficients, are also needed to be explained. The determinants of the housing price in the central area is more complex than other areas and this may mislead the common impact of distance to park on housing price. For the Zhengdong New District, the opposite coefficients may because most parks are small in size and lack sufficient appeal. Also, most of the newly built communities have equipped with a good green space and some fitness facilities, which offset the impact on the surrounding

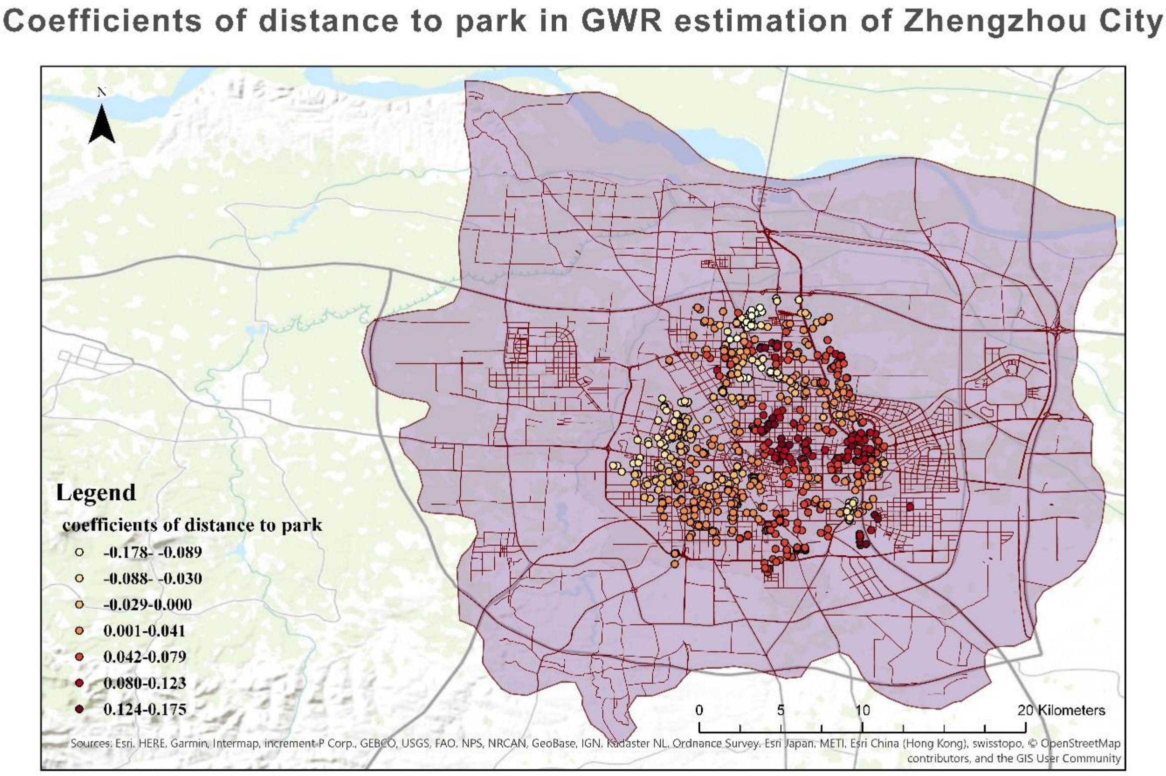

Figure 25 Distribution of coefficients of distance to park in GWR estimation

housing prices to some extent.

(8) Distance to hospital

The coefficients of distance to nearby hospital vary from -0.159 to 0.197, and it is significant at a 0.1 significance level. As shown in Figure 26, the areas with different colours represent different coefficients of distance to hospital based on the inverse distance weighted (IDW) technique in ArcGIS. Besides, the blue dots show the distribution of hospitals in Zhengzhou City. Most of the areas have negative coefficients which indicates that located near a hospital is an advantage of a house, which is similar to what has been expected based on the literature review. However, there is a clear clue finding that the areas with positive coefficients, corresponding to the orange and red parts, have higher housing prices with the distance to the nearby hospitals increasing. This is inconsistent with the theoretical expectation. The possible explanations are as follows: since the density of hospital in these areas is too low to meet the daily need for the nearby residents, it causes the distortion of the influence of hospital on housing price. Moreover, instead of general hospitals, many hospitals located there are specialized hospitals. As a result, residents have to choose hospitals with corresponding functions according to their condition and concern little on the distance.

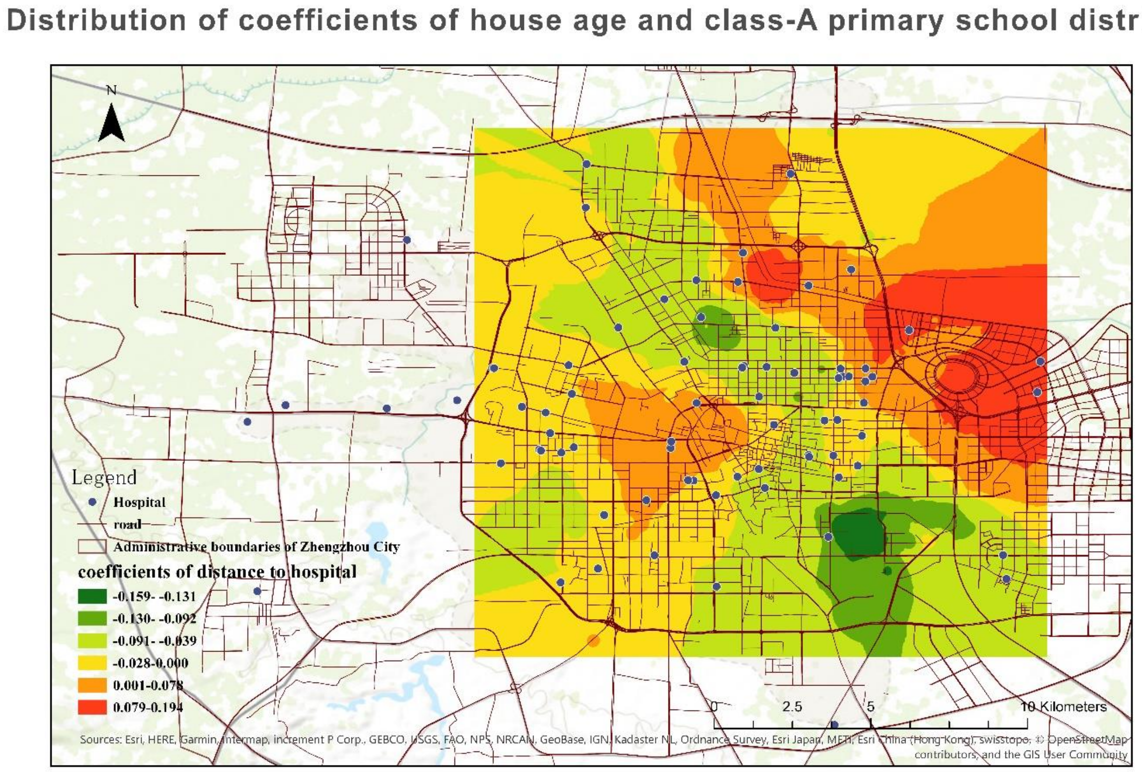

Figure 26 Distribution of coefficients of distance to hospital in GWR estimation

(9) Distance to subway station.

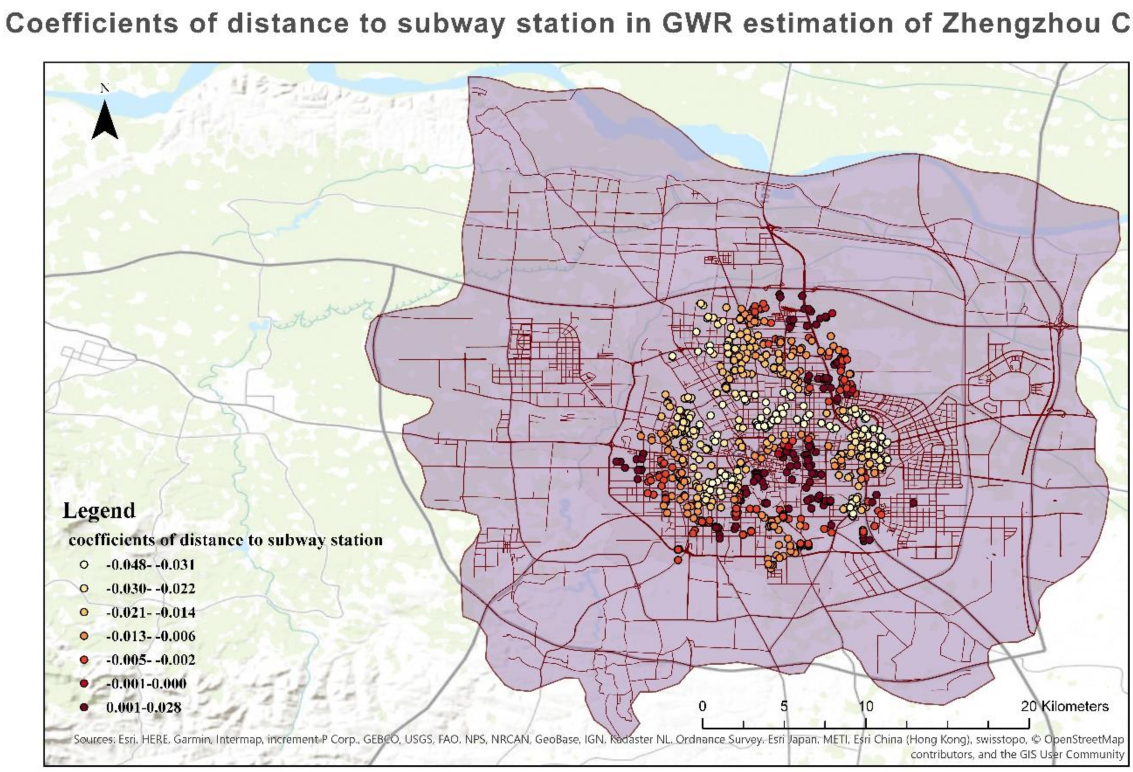

Figure 27 Distribution of coefficients of distance to hospital in GWR estimation

According to Table 7, the coefficients of distance to subway station range from -0.048 to 0.028. 83% of the coefficients of distance to subway station are negative, implying this indicator generally has a negative impact on housing price. In specific, living near a subway station will increase the overall housing price. From Figure 27, the dark red points (occupied 17% of the whole sample points) have positive coefficients, this may happen because people don't like the noise brought by the nearby substation. For those areas with negative values, the houses in periphery areas tend to have a stronger rely on the subway station. Since these areas are far from the central city, it takes a long time to travel by public transport. The opening of the subway has greatly improved the traffic conditions in these areas, therefore, the residential price is sensitive to the distance of the subway station. The value of the regression coefficient in the inner city is relatively small. The main reason is that the main roads there are densely distributed, people are convenient to travel, and the dependence on the subway is relatively small. Therefore, the impact of the subway on the residential price is not obvious.

(10) Distance to main road

The overall range of the coefficients of distance to main road is from -0.418 to 0.263. As shown obvious in Figure 28, the coefficients in the outskirt tend to be positive and those in the central city are negative. It reveals that people, living in the inner city, concern more of the distance to main road and prefer to live nearby a main road for a better traffic convenience. Consequently, distance to the main road poses a negative impact on the housing price. For those living in the periphery areas of Zhengzhou City, the positive influence of this indicator can be explained like, people prefer to live in a quite environment and don't want to be bordered by the noise. As a result, the longer distance from the main road, the higher the housing price.

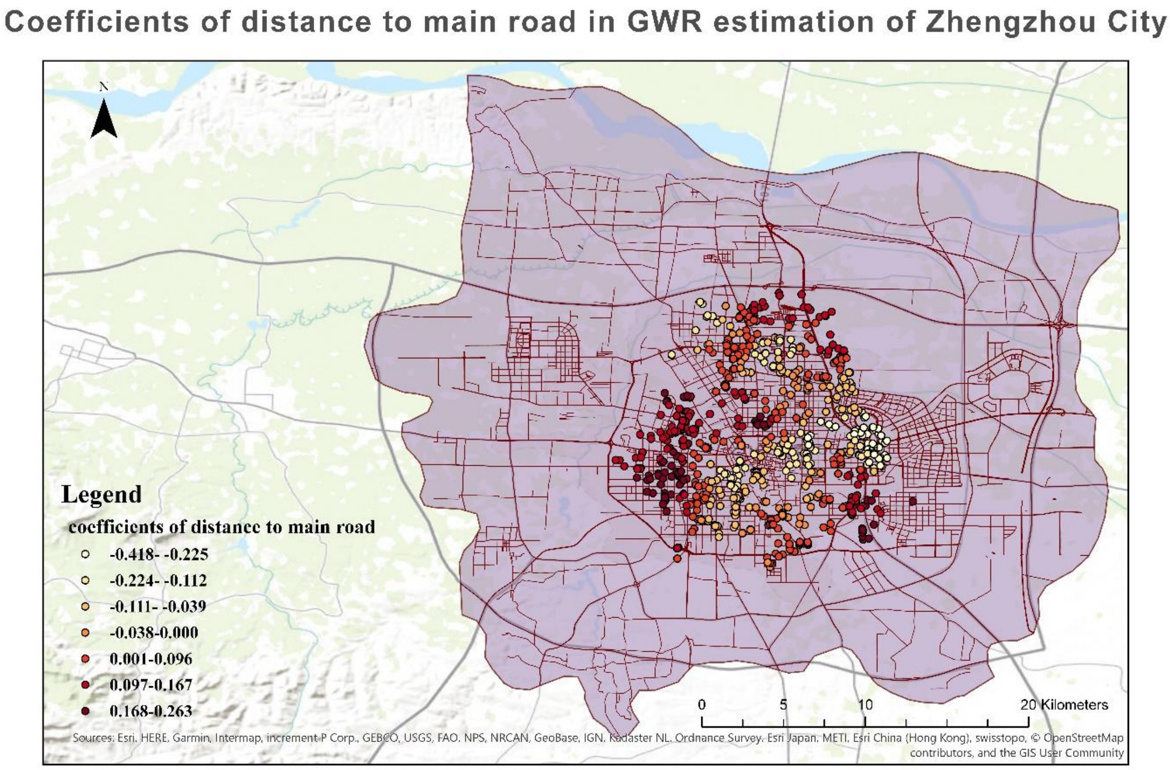

Figure 28 Distribution of coefficients of distance to subway station in GWR estimation

(11) Distance to CBD (Erqi Square)

The coefficients of distance to CBD range between -1.974 and 3.618, and it is significant at a 0.1 significance level. The below Figure 29 shows a quite large area of positive coefficient of distance to CBD distributed in Zhengzhou City. It somehow indicates that people in Zhengzhou City don’t like to live nearby the CBD. This may because of the noise in CBD on one hand, and also could be a signal of the suburbanization phenomenon in Zhengzhou City, especially the dark red dots with the highest positive coefficients.

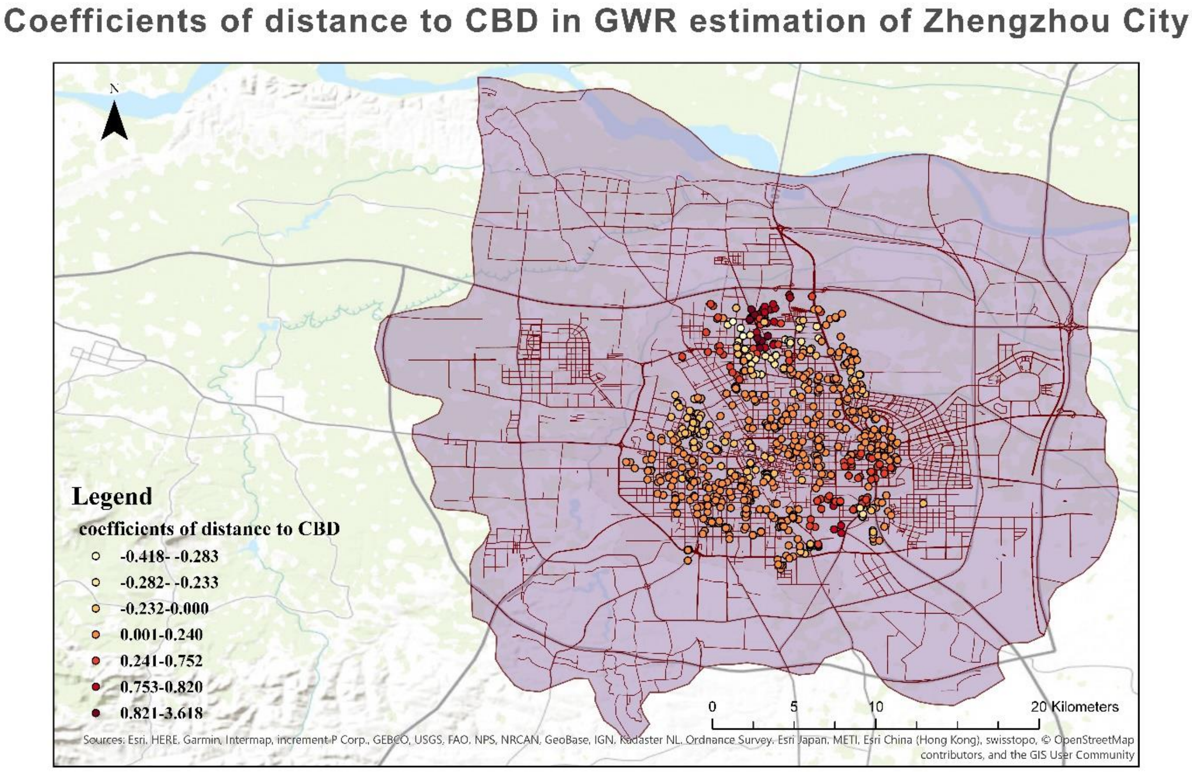

Figure 29 Distribution of coefficients of distance to CBD in GWR estimation

(12) Distance to Jinshui River Park

Coefficients of distance to Jinshui River Park fill in the interval between -3.648 and 2.071. Among them, 82% of the sample points identify a negative effect of this indicator on housing price, which implies that houses located near the Jinshui River Park will deserve a higher price. Jinshui River is the mother river of Zhengzhou City, with beautiful scenery and a relaxing environment. Consequently, as what has been expected, Jinshui River Park has an obvious positive effect on housing price. Those who show a negative coefficient, which is located quite far away from the Jinshui River Park (Figure 30), can be explained as: because of the far distance, the effect of Jinshui River Park cannot reach those houses. As a result, Jinshui River Park has an opposite effect on housing price of them.

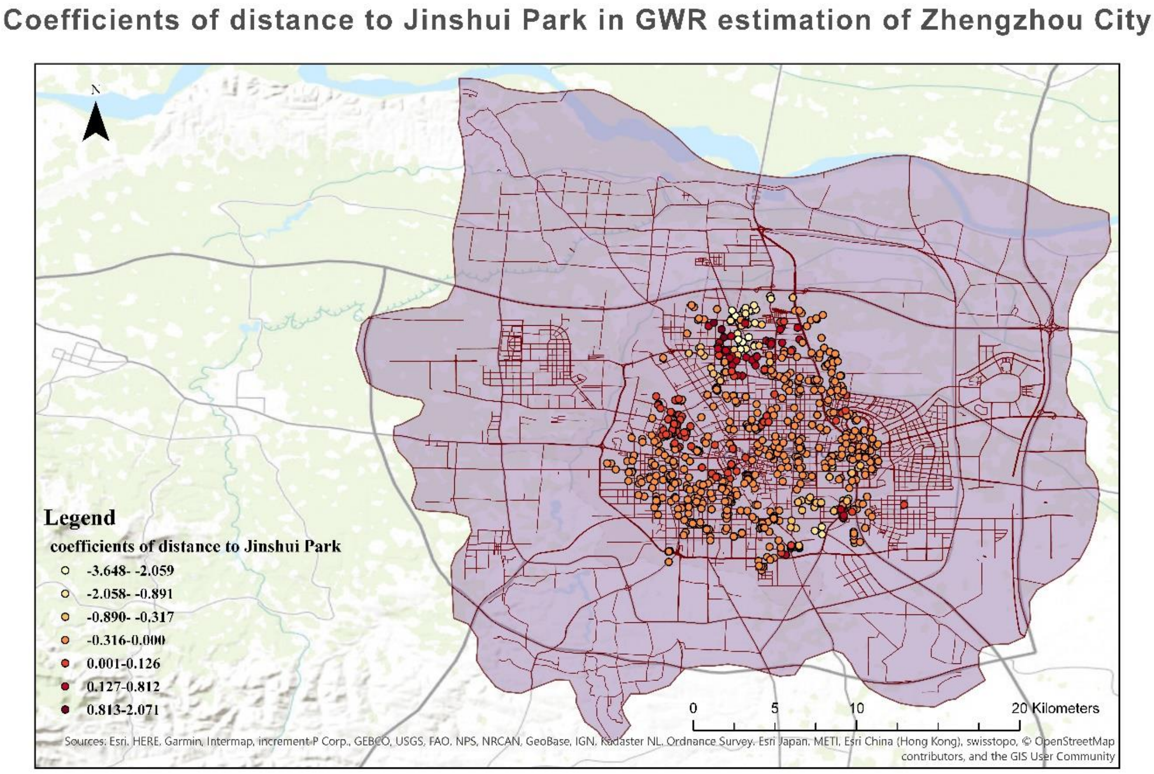

Figure 30 Distribution of coefficients of distance to Jinshui River Park in GWR estimation

4.4 Model Comparison and Summary

Commonly, the comparison between the two models based on OLS and GWR estimations are generated by the following four aspects: the improvement of adjusted R-squared; the reduction of AICc value (at least 3 as noted above); the reduction of Moran's I which measures the randomness of spatial autocorrelation of residuals; ANOVA test of the residuals of the two models (Mashhoodi, 2018).

In this case, comparison between these two models are listed in Table 8. Firstly, having a 6% improvement of the adjusted R-squared, the GWR estimation explains more of the variation of housing price than the OLS estimation. Also, the range of local R^2 ranges from 0.450 to 0.822 (see Figure 18), and it indicates that 79% of the sample points has a higher goodness-of-fit than the same points in OLS estimation. And the geographical map delineates that, except the points located in a small area of northern city, other sample dots have relative higher values of R^2. This also emphasizes that the GWR estimation is better than the OLS estimation. Secondly, the AICc of the GWR estimation is -6263, much lower than that of OLS estimation (-5298). Thirdly, the Moran's Index for the residuals in the OLS estimation is 68.756, while the number is shrunk to 8.5 in the GWR estimation (see Figure 31), indicating a better performance of GWR estimation. Although the residuals are still distributed clustered, the z-score has already been strongly reduced closing to 0. Lastly, the result of ANOVA test is reported in Table 9. The residuals of GWR estimation has been reduced by 33.480, compared with that of OLS estimation. In brief, the GWR estimation has a better performance when analyzing the impact of housing attributes on housing price than the OLS estimation.

Table 8 Estimated parameters and diagnostic statistics in the OLS and GWR estimations

Variable	GWR results				OLS results
	β Mean	β Min	β Max	β SD	β
Intercept	9.707^{***}	6.194	11.536	0.573	9.546^{***}
Structural characteristics					
lgsize	0.885^{***}	0.789	0.998	0.045	0.901^{***}
Floors	0.001	-0.002	0.003	0.001	0.001^{***}
Building type	0.019	-0.077	0.080	0.026	0.044^{***}
Building orientation	0.057^{*}	-0.040	0.135	0.038	0.064^{***}
House age	-0.012^{***}	-0.024	0.004	0.006	-0.010^{***}
Neighborhood characteristics					
Plot ratio	-0.006^{*}	-0.045	0.029	0.017	0.001
Green ratio	0.366^{***}	-0.328	0.811	0.229	0.457^{***}
DIST park	0.028^{*}	-0.178	0.175	0.061	0.007^{**}
DIST hospital	-0.014^{*}	-0.159	0.197	0.058	0.015^{***}
School district	-0.007	-0.255	0.180	0.080	0.029^{***}
Locational characteristics					
Number of bus stops	0.005	-0.024	0.044	0.013	0.007^{***}
DIST subway station	-0.012^{**}	-0.048	0.028	0.013	-0.012^{***}
DIST main road	0.021^{*}	-0.418	0.263	0.137	-0.018
DIST CBD	0.132^{*}	-1.974	3.618	0.508	0.034^{***}
DIST Jinshui River Park	-0.149^{*}	-3.648	2.071	0.513	-0.051^{***}
R-squared	**0.697**				**0.627**
Adjusted R-squared	**0.684**				**0.626**
AICc	**-6262.754**				**-5298.056**
Residuals Moran's I	**8.545**				**68.756**

β : * significane at the 0.000 level, ** significane at the 0.05 level, * significane at the 0.1 level.**

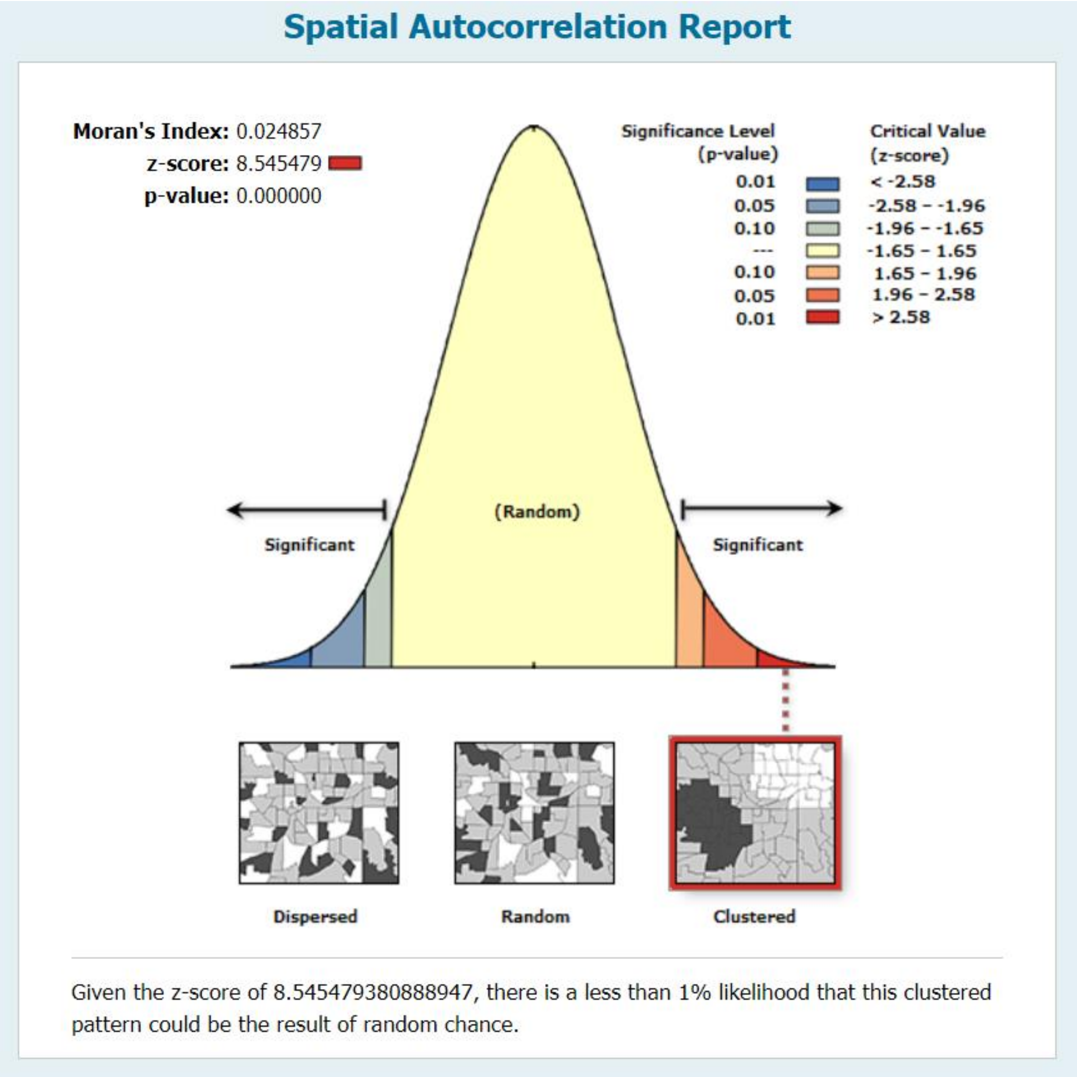

Figure 31 Spatial autocorrelation report of the residuals of GWR estimation

Table 9 ANOVA table of the residuals in OLS and GWR estimations

Source	Sum Sq	DF	Mean Sq	F value
OLS Residuals	176.240	6668.000		
GWR Improvement	33.480	273.949	0.122	
GWR Residuals	142.760	6394.051	0.022	5.474***

***** p-value < 0.01.**

4.5 Summary

This chapter conducts the regression estimation of the OLS estimation and the GWR estimation for analyzing the impact of housing attributes on housing price, and compares them by the adjusted R^2, AICc, Moran's Index and the ANOVA test for residuals. The relative comments and conclusions are as follows:

(1) According to the sample data of the study area, the descriptive statistics of the variables are carried out, and it is found that the housing price and each explanatory variable have great variations. The correlation among variables are calculated, and the correlation analysis is reported. The results obtained by constructing an OLS estimation show that the VIF values of

the independent variables are less than 7.5, indicating that there is no collinearity between the variables. The R^2 of the model is 0.626, implying the goodness-of-fit of the model is quite good. However, there is an aggregation phenomenon of the residuals, indicating that the ordinary linear regression model has bias and is not suitable for analyzing the influencing factors of residential prices.

(2) Based on the heterogeneity of the distribution of housing price, spatial coordinate information is introduced to conduct the spatial autocorrelation and agglomeration analysis. After that, a GWR estimation is built to analyze the correlation between housing attributes and housing price. Through the fitting analysis of the model and the significance test of the overall regression relationship, compared with the OLS estimation, the adjusted R^2 of the GWR estimation is increased from 0.626 to 0.684. Besides, the sum of the squared residual value is reduced by 33.48. The spatial autocorrelation of the residual, expressed by the Global Moran's Index is greatly decreased to 8.5, indicating the accuracy of the regression estimation has been significantly improved. The geo maps of the spatial residual distribution and the corresponding regression coefficients of the significant independent variables are provided. These all demonstrate that, taking the spatial heterogeneity of housing price into account, the GWR estimation is more suitable for analyzing the relationship between the influencing factors and housing price in Zhengzhou City.

Using the ArcGIS software, the statistic results of the GWR estimation and the variation of regression coefficients of each dependent indicator can be visually displayed in a series of geo-maps. All in all, the determinants of housing price in Zhengzhou City can be concluded as follows: Being significant at a minimum of 0.1 significance level, the building size, building orientation, green ratio, distance to park, distance to main road, distance to CBD have positive effect on housing price, while the house age, plot ratio, distance to hospital, distance to subway station, distance to Jinshui River Park pose negative impact on residential price.

Chapter 5: Conclusions and recommendations

5.1 Retrospect

As what has been shown in the first chapter, the residential housing market has grown substantially over the last twenty years in China. Generally, housing price is a hot topic in China and is affected by factors such as macroeconomics, market supply and demand, and national policies. It is also closely related to micro factors such as geographic location and housing characteristics. The analysis and forecast of urban housing prices have important significance and practical value for housing security, real estate development and urban planning. However, the existing body of literature implicitly presumed that the influence of each housing characteristic on housing price is stationary, which is obviously not completely consistent with reality and is an unrealistic simplification of residential dwelling markets. Therefore, this paper was aimed at fixing the knowledge gap and analyzing the correlation between housing attributes and housing price, taking the spatial heterogeneity into account. By conducting a model based on the Geographically Weighted Regression (GWR) estimation technique, this paper analysed the spatial distribution of second-hand residential houses in Zhengzhou City and revealed the mechanism of various determinants impacting on houses.

In specific, this research focused on analyzing the determinants of second-hand housing prices in Zhengzhou City incorporating spatial heterogeneity. The housing transaction samples in 2016 were selected and various models using OLS and GWR estimation techniques were used in this thesis to detect the relationship between housing attributes and housing prices in Zhengzhou City. These two approaches were conducted for one main research question and two sub-questions respectively.

The main question is:

What are the determinants of second-hand housing prices and do these determinants vary between different areas?

The sub-questions are:

a. Based on Hedonic Price Model, what factors and how they will influence the second-hand housing prices in Zhengzhou City?

b. Is there spatial heterogeneity in the price composition of the second-hand housing market in Zhengzhou City?

5.2 Conclusions

This paper takes the transaction housing price data in Zhengzhou City in 2016 as the research object, and uses the Exploratory Spatial Data Analysis (ESDA) method in ArcGIS to detect the distribution of housing price and its development trend. Based on models using OLS estimation and GWR estimation, the relationship is built between the housing characteristics and housing price. The following conclusions are drawn:

(1) By taking advantage of the Exploratory Spatial Data Analysis (ESDA), including the Normal Distribution Test, Trend Analysis and Spatial Autocorrelation Analysis, some spatial features can be detected from the corresponding analysis. To be exact, the trend analysis function of the geostatistical analysis in ArcGIS could provide a spatial trend chart of the distribution of housing price in Zhengzhou City. Figure 9 and Figure 10 show that the second-hand housing price in Zhengzhou has a distinct monocentric structure in both the north and south, east and west directions, and the downward spiral from the center to the edge is very obvious. The decreasing range of the north-south direction and the east-

west direction are roughly the same. This indicates that the construction of Zhengzhou City is radiated from the center of the Erqi Square (CBD), and gradually forms sub-centres in some areas, especially in the east and in the northwest. Among these sub-centres, the development of Zhengdong New District is the fastest, and housing price here is likely to catch up with the that in the CBD in the near future. Also, the overall housing price in Zhengzhou City tends to distributed clustered based on the Global Moran's Index and has already formed high-high clusters in CBD and Zhengdong New District based according to the Local Moran's Index.

(2) The OLS estimation can explain the reasons for the change of housing price to a certain extent (0.625), but there is a large spatial agglomeration of the residuals, which implies that the changes are non-stationary and needed to be tested in a GWR estimation. By comparing the Hedonic Price Model constructed by least squares method and geographic weighted regression method, it is found that the GWR estimation has a better goodness-of-fit (0.684). Through the local analysis of the independent variables, the GWR estimation can better reflect the spatial heterogeneity of the influencing factors, and can analyze each factor separately through visualization methods in ArcGIS to reveal more characteristics and phenomena that cannot be reflected by an OLS estimation. Although the residuals in GWR estimation are still not distributed randomly, the value has been decreased from 68 (OLS estimation) to 8.5. All in all, the use of GWR estimation can better explain the correlation between housing attributes and second-hand housing prices in Zhengzhou City.

(3) Through the visual analysis of the GWR estimation of local R^2, it is found that the local R^2 in the northwest part of Zhengzhou City is lower than other areas, which may be the secondary core of the urban housing structure. This finding is different from that in the previous studies, Zhengdong New District is the only sub-centre of Zhengzhou City. This conclusion can be studied in the subsequent researches.

(4) According to the results of the Hedonic Price Model constructed by the GWR estimation in GWR 4.0, 11 housing attributes have significant impact on the second-hand housing price in Zhengzhou City. They are the building size, building orientation, house age, plot ratio, green ratio, distance to park, distance to hospital, distance to subway, distance to main road, distance to CBD and distance to Jinshui River Park. In terms of the structural characteristics, the building size and the building orientation have a positive influence on the housing price. While the house age poses a negative impact on the residential price. For the neighborhood characteristics, the green ratio and distance to park exert a positive influence on the housing value while the plot ratio and the distance to hospital influenced housing price negatively. According to the locational characteristics, the houses locates near a subway station and the Jinshui River Park have higher price. Whereas, those stand near the main roads and CBD usually have lower housing price. The weight of these influencing factors shows a trend of change in space in GWR estimation with their different spatial locations and reflects the spatial heterogeneity of the influence of these indicators. These all show that, taking the spatial heterogeneity of housing price into account, the GWR estimation is more suitable for analyzing the relationship between the influencing factors and housing price in Zhengzhou City.

5.3 Recommendations

Corresponding to the objective of this research, three recommendations can be concluded to the future real estate appraisal method:

(1) Governments may consider using Hedonic Price Modelling as a basis for levying real estate tax, instead of the Land Datum Value Method or the Market Comparison Approach,

as hedonic modelling provides more accurate results (see conclusions). Especially in China, where real estate taxation is a relatively new phenomenon, and thus there is an opportunity to start 'fresh', using Computer assisted Mass Appraisal Models based on the Hedonic Price Model can ensure fair and accurate tax collection.

(2) Incorporate spatial heterogeneity in appraisal models. By including some form of spatial weighting or estimation (such as GWR), the basis for tax collection can be improved. It is possible to include such an estimation technique in Computer assisted Mass Appraisal models. However, further study may be needed to define the exact spatial parameters.

(3) Explicitly model submarket in Computer Assisted Mass Appraisal Models. By separately modelling different submarkets, for instance segmented on price, type or even resident it is possible to create more accurate pricing models, which further improves the accuracy of real estate tax collection.

5.4 Limitations and prospect

Due to the limited level of knowledge and time, there are still many limitations in this study that can be improved, and the research content can be further deepened.

The limitations and prospect of this study could be analysed as follows:

(1) In this paper, because of the unavailability of the panel data, this study of second-hand housing prices only takes the spatial characteristics into consideration and did not go deep into studying the housing price changes over time. Actually, the change of the second-hand housing price in Zhengzhou over time is also very important for the formulation of public policies and the guidance of second-hand housing transactions. Besides, the endogeneity problem caused by cross-sectional data which refers to the circulated impact of the spillover effect of housing price on spatial heterogeneous coefficients of different indicators is another limitation of this research. In a nutshell, the relationship between housing attributes and second-hand housing prices can be further analysed in combination with spatial heterogeneity and time factors.

(2) According to the Hedonic Price Model constructed by the geographical weighted regression method in this paper, the existing sub-center of Zhengdong New District and the potential secondary core of North Longhu District are parts of the urban spatial structure of Zhengzhou City. However, this paper mainly treats the city center as the core. As a result, on the basis of the research in this paper, the consideration of several secondary cores in the future researches can improve the goodness-of-fit of the model, and can further deepen the study of the housing influencing factors incorporating spatial heterogeneity.

(3) Due to the limited processing power of GWR 4.0 software, only 6684 out of 10524 observations could be analysed in this research. In the future study, it is necessary to test these sample points in other spatial software.

Bibliography

Ahn, J. J., Byun, H. W., Oh, K. J. and Kim, T. Y. 2012. Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications,* 39 (9), pp. 8369-8379. [Accessed 4/17/2018 5:55:00 AM].

Anselin, L., 1999. Interactive techniques and exploratory spatial data analysis. *Geographical Information Systems: Principles, Techniques, Management and Applications,* 1 pp. 251-264. [Accessed 7/26/2018 4:29:06 AM].

Arbia, G., 2001. The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems,* 3 (3), pp. 271-281. [Accessed 7/26/2018 5:08:39 AM].

Baldassare, M., Knight, R. and Swan, S. 1979. Urban service and environmental stressor: The impact of the Bay Area Rapid Transit System (BART) on residential mobility. *Environment and Behavior,* 11 (4), pp. 435-450. [Accessed 4/16/2018 3:14:24 PM].

Ball, M. J., 1973. Recent empirical work on the determinants of relative house prices. *Urban Studies,* 10 (2), pp. 213-233. [Accessed 4/15/2018 8:17:46 AM].

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. [Accessed 4/15/2018 9:25:05 AM].

Boarnet, M. G., 1994. The monocentric model and employment location. *Journal of Urban Economics,* 36 (1), pp. 79-97. [Accessed 4/16/2018 2:29:01 PM].

Bourassa, S., Cantoni, E. and Hoesli, M. 2010. Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research,* 32 (2), pp. 139-159. [Accessed 4/17/2018 5:55:30 AM].

Brasington, D. M. and Hite, D. 2005. Demand for environmental quality: a spatial hedonic analysis. *Regional Science and Urban Economics,* 35 (1), pp. 57-82. [Accessed 4/13/2018 4:55:43 PM].

Brunsdon, C., Fotheringham, A. and Charlton, M. 2002. Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems,* 26 (6), pp. 501-524. [Accessed 4/18/2018 8:38:44 AM].

Butler, R. V., 1982. The specification of hedonic indexes for urban housing. *Land Economics,* 58 (1), pp. 96-108. [Accessed 4/15/2018 3:13:19 AM].

Carroll, T. M., Clauretie, T. M. and Jensen, J. 1996. Living next to godliness: Residential property values and churches. *The Journal of Real Estate Finance and Economics,* 12 (3), pp. 319-330. [Accessed 4/16/2018 4:14:39 AM].

Cassel, E. and Mendelsohn, R. 1985. The choice of functional forms for hedonic price equations: comment. *Journal of Urban Economics,* 18 (2), pp. 135-142. [Accessed 4/15/2018 4:47:35 AM].

Chau, K., Ma, V. and Ho, D. 2001. The pricing of 'luckiness' in the apartment market. *Journal of Real Estate Literature,* 9 (1), pp. 29-40. [Accessed 4/15/2018 8:18:41 AM].

Chau, K. W. and Chin, T. 2002. A critical review of literature on the hedonic price model. [Accessed 4/14/2018 8:29:28 AM].

Chen, B., 2009. The Research of Housing Characteristic Pric in Guangzhou Based on Hedonic Model. *Master Thesis, Jinan University,* [Accessed 4/14/2018 9:20:18 AM].

Clapp, J. M. and Giaccotto, C. 1998. Residential hedonic models: A rational expectations approach to age effects. *Journal of Urban Economics,* 44 (3), pp. 415-437. [Accessed 4/16/2018 5:16:21 AM].

Clark, D. E. and Herrin, W. E. 2000. The impact of public school attributes on home sale prices in California. *Growth and Change,* 31 (3), pp. 385-407. [Accessed 4/16/2018 10:07:53 AM].

Clauretie, T. M. and Neill, H. R. 2000. Year-round school schedules and residential property values. *The Journal of Real Estate Finance and Economics,* 20 (3), pp. 311-322. [Accessed 4/16/2018 10:08:22 AM].

Court, A., 1939. The dynamics of automobile demand. *The Dynamics of Automobile Demand,* [Accessed 4/15/2018 4:34:54 PM].

Cox, N., 2006. WINSOR: Stata module to Winsorize a variable. [Accessed 7/27/2018 3:33:38 AM].

Dong, K., 2017. Study on the evaluation of housing price in Zhengzhou based on hedonic price model . *Shihezi University,* .

Dubin, R. A., 1988. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics,* pp. 466-474. [Accessed 4/13/2018 4:53:18 PM].

Dunse, N. and Jones, C. 1998. A hedonic price model of office rents. *Journal of Property Valuation and Investment,* 16 (3), pp. 297-312. [Accessed 4/15/2018 8:08:52 AM].

Espey, M. and Lopez, H. 2000. The impact of airport noise and proximity on residential property values. *Growth and Change,* 31 (3), pp. 408-419. [Accessed 4/16/2018 10:11:22 AM].

Farber, S. and Yeates, M. 2006. A comparison of localized regression models in a hedonic house price context. *Canadian Journal of Regional Science,* 29 (3), pp. 405-420. [Accessed 4/18/2018 5:52:01 AM].

Fik, T. J., Ling, D. C. and Mulligan, G. F. 2003. Modeling spatial variation in housing prices: a variable interaction approach. *Real Estate Economics,* 31 (4), pp. 623-646. [Accessed 4/18/2018 5:50:47 AM].

Follain, J. R. and Jimenez, E. 1985. Estimating the demand for housing characteristics: a survey and critique. *Regional Science and Urban Economics,* 15 (1), pp. 77-107. [Accessed 4/15/2018 10:36:35 AM].

Forrest, D., Glen, J. and Ward, R. 1996. The impact of a light rail system on the structure of house prices: a hedonic longitudinal study. *Journal of Transport Economics and Policy,* pp. 15-29. [Accessed 4/16/2018 5:22:07 AM].

Fotheringham, A. S., Brunsdon, C. and Charlton, M., 2000. Quantitative geography: perspectives on spatial data analysis. Sage. [Accessed 5/6/2018 10:07:20 AM].

Freeman, A. M., 1981. Hedonic prices, property values and measuring environmental benefits: a survey of the issues. Hedonic prices, property values and measuring environmental benefits: a survey of the issues. 1981. Measurement in public choice. Springer. pp. 13-32. [Accessed 4/15/2018 8:16:45 AM].

Garrod, G. D. and Willis, K. G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. *Journal of Environmental Management,* 34 (1), pp. 59-76. [Accessed 4/16/2018 6:00:10 AM].

Gillard, Q., 1981. The effect of environmental amenities on house values: the example of a view lot. *The Professional Geographer,* 33 (2), pp. 216-220. [Accessed 4/16/2018 4:20:07 AM].

Goodman, A. C., 1978. Hedonic prices, price indices and housing markets. *Journal of Urban Economics,* 5 (4), pp. 471-484. [Accessed 4/16/2018 6:28:37 AM].

Griliches, Z., 1961. Hedonic price indexes for automobiles: An econometric of quality change. Hedonic price indexes for automobiles: An econometric of quality change. 1961. The Price Statistics of the Federal Goverment. NBER. pp. 173-196. [Accessed 4/15/2018 4:35:52 PM].

Guillain, R. and Le Gallo, J. 2010. Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. *Environment and Planning B: Planning and Design,* 37 (6), pp. 961-981. [Accessed 7/26/2018 5:08:08 AM].

Halvorsen, R. and Pollakowski, H. O. 1981. Choice of functional form for hedonic price equations. *Journal of Urban Economics,* 10 (1), pp. 37-49. [Accessed 4/15/2018 3:14:16 AM].

Haurin, D. R. and Brasington, D. 1996. School quality and real house prices: Inter-and intrametropolitan effects. *Journal of Housing Economics,* 5 (4), pp. 351-368. [Accessed 4/16/2018 3:42:23 PM].

Healy, M. and Perry, C. 2000. Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal,* 3 (3), pp. 118-126. [Accessed 5/10/2018 3:21:49 AM].

Helbich, M., Brunauer, W., Hagenauer, J. and Leitner, M. 2013. Data-driven regionalization of housing markets. *Annals of the Association of American Geographers,* 103 (4), pp. 871-889. [Accessed 4/17/2018 5:47:20 AM].

Helbich, M. and Griffith, D. A. 2016. Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems,* 57 pp. 1-11. [Accessed 4/17/2018 5:44:51 AM].

Helbich, M. and Jokar Arsanjani, J. 2015. Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science,* 42 (2), pp. 134-148. [Accessed 4/18/2018 5:51:27 AM].

Huh, S. and Kwak, S. 1997. The choice of functional form and variables in the hedonic price model in Seoul. *Urban Studies,* 34 (7), pp. 989-998. [Accessed 4/16/2018 10:09:49 AM].

Jensen, J., Liljemark, W. and Bloomquist, C. 1981. The effect of female sex hormones on subgingival plaque. *Journal of Periodontology,* 52 (10), pp. 599-602. [Accessed 4/15/2018 2:58:34 AM].

Ketkar, K., 1992. Hazardous waste sites and property values in the state of New Jersey. *Applied Economics,* 24 (6), pp. 647-659. [Accessed 4/16/2018 10:05:42 AM].

Kohlhase, J. E., 1991. The impact of toxic waste sites on housing values. *Journal of Urban Economics,* 30 (1), pp. 1-26. [Accessed 4/16/2018 3:29:13 PM].

Lancaster, K. J., 1966. A new approach to consumer theory. *Journal of Political Economy,* 74 (2), pp. 132-157. [Accessed 4/15/2018 4:46:29 PM].

Leggett, C. G. and Bockstael, N. E. 2000. Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management,* 39 (2), pp. 121-144. [Accessed 4/15/2018 8:19:21 AM].

LeSage, J. P. and Pace, R. K. 2004. Models for spatially dependent missing data. *The Journal of Real Estate Finance and Economics,* 29 (2), pp. 233-254. [Accessed 4/16/2018 2:47:12 AM].

Li, M. M. and Brown, H. J. 1980. Micro-neighborhood externalities and hedonic housing prices. *Land Economics,* 56 (2), pp. 125-141. [Accessed 4/16/2018 4:17:55 AM].

Li, Z. and Wu, F. 2008. Tenure-based residential segregation in post-reform Chinese cities: a case study of Shanghai. *Transactions of the Institute of British Geographers,* 33 (3), pp. 404-419. [Accessed 4/4/2018 4:05:17 PM].

Linneman, P., 1980. Some empirical results on the nature of the hedonic price function for the urban housing market. *Journal of Urban Economics,* 8 (1), pp. 47-68. [Accessed 4/15/2018 4:53:04 AM].

Liu, J., 2016. Study on the Impact of House Characteristics on the Asking Price of Urban Second-hand House — a Case of Xuzhou City . *Master Thesis, China Mining University,* [Accessed 4/18/2018 8:41:18 AM].

Liu, K., Wu, Q. and Wang, P. 2015. Econometric analysis of the impacts of rail transit on property values： the number 1 and 2 lines in Nanjing. *Resources Science,* 37 (1), pp. 133-141. [Accessed 4/16/2018 2:55:46 PM].

Liu, X., 2017. Study on determinants of housing price in Kunming based on spatial econometrical models. *Master Thesis, Yunnan University of Finance and Economics,* [Accessed 4/17/2018 1:06:53 PM].

Lu, B., Charlton, M., Harris, P. and Fotheringham, A. S. 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science,* 28 (4), pp. 660-681. [Accessed 4/17/2018 5:47:58 AM].

Malpezzi, S., 2003. Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy,* pp. 67-89. [Accessed 4/14/2018 4:42:05 AM].

Mashhoodi, B., 2018. Spatial dynamics of household energy consumption and local drivers in Randstad, Netherlands. *Applied Geography,* 91 pp. 123-130. [Accessed 8/6/2018 8:46:02 AM].

McCord, M., Davis, P., Haran, M., McGreal, S., 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. [Accessed 4/15/2018 9:27:09 AM; 4/15/2018 9:27:09 AM].

McMillen, D. P. and Redfearn, C. 2010. Estimation, interpretation, and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science,* 50 (3), pp. 712-733. [Accessed 4/17/2018 5:52:55 AM].

Ming, K. and Hou, X. 2017. Research on the Influence of Housing Characteristics on Real Estate Value Based on Hedonic Price Model. *Shangyejingji,* 12 pp. 042. [Accessed 4/16/2018 2:08:41 PM].

Moran, P. A., 1950. Notes on continuous stochastic phenomena. *Biometrika,* 37 (1/2), pp. 17-23. [Accessed 8/27/2018 4:28:12 AM].

Osland, L., 2010. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research,* 32 (3), pp. 289-320. [Accessed 4/18/2018 3:50:20 AM].

O'sullivan, D. and Unwin, D., 2014. Geographic information analysis. John Wiley & Sons. [Accessed 7/27/2018 8:30:49 AM].

Pace, R. K., Barry, R. and Sirmans, C. 1998. Spatial statistics and real estate. *The Journal of Real Estate Finance and Economics,* 17 (1), pp. 5-13. [Accessed 4/18/2018 5:50:14 AM].

Pace, R. K. and LeSage, J. P. 2010. Omitted variable biases of OLS and spatial lag models. Omitted variable biases of OLS and spatial lag models. 2010. Progress in Spatial Analysis. Springer. pp. 17-28. [Accessed 4/13/2018 4:51:07 PM].

Paez, A., Uchida, T. and Miyamoto, K. 2001. Spatial association and heterogeneity issues in land price models. *Urban Studies,* 38 (9), pp. 1493-1508. [Accessed 4/17/2018 5:51:58 AM].

Palmquist, R. B., 1992. Valuing localized externalities. *Journal of Urban Economics,* 31 (1), pp. 59-68. [Accessed 4/16/2018 10:10:56 AM].

Peterson, S. and Flanagan, A. 2009. Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research,* 31 (2), pp. 147-164. [Accessed 4/15/2018 4:58:35 PM].

Rasmussen, D. W. and Zuehlke, T. W. 1990. On the choice of functional form for hedonic price functions. *Applied Economics,* 22 (4), pp. 431-438. [Accessed 4/15/2018 4:32:40 AM].

Richardson, H. W., Vipond, J. and Furbey, R. A. 1974. Determinants of urban house prices. *Urban Studies,* 11 (2), pp. 189-199. [Accessed 4/16/2018 9:44:24 AM].

Rodriguez, M. and Sirmans, C. 1994. Quantifying the value of a view in single-family housing markets. *Appraisal Journal,* 62 pp. 600-600. [Accessed 4/16/2018 5:19:16 AM].

Rong, L., 2014. Spatial Distribution and Reason Analysis of Second-hand House Price Based on GIS and Hedonic Model--A Case of Kunming City. *Master Thesis, the Kunming University of Science and Technology,* [Accessed 4/5/2018 2:49:40 PM].

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy,* 82 (1), pp. 34-55. [Accessed 4/15/2018 4:36:45 PM].

She, M., 2017. Research on the Spatial Variation of Hedonic Prices of Hangzhou Office Rent Price Based on Geographically Weighted Regression Model. *Zhejiang University,* [Accessed 5/14/2018 3:33:45 AM].

Shen, X., 2015. Problems and Suggestions of the Secondary-hand Housing Market. *Marketing Research,* (12), pp. 7-9. [Accessed 4/15/2018 4:16:41 PM].

Si, Z. and Shi, Y. 2013. The Impacts of Class 3A Comprehensive Hospitals on Population Distribution and Housing Price—Take Children's Hospital of Fudan University for Example. *Economic Geography,* 33 (10), pp. 74-81. [Accessed 4/16/2018 10:36:33 AM].

Sirmans, S., Macpherson, D. and Zietz, E. 2005. The composition of hedonic pricing models. *Journal of Real Estate Literature,* 13 (1), pp. 1-44. [Accessed 5/8/2018 2:19:08 AM].

So, H., Tse, R. Y. and Ganesan, S. 1997. Estimating the influence of transport on house prices: evidence from Hong Kong. *Journal of Property Valuation and Investment,* 15 (1), pp. 40-47. [Accessed 4/15/2018 4:52:03 AM].

Straszheim, M. R., 1975. An econometric analysis of the urban housing market. *NBER Books,* [Accessed 4/18/2018 3:51:24 AM].

Su, Y., Zhu, D., Zheng, Y., Wang, X., et al., 2015. The effects of subways on housing price gradients between urban area and suburb in southwest Beijing . *Resources Science,* 37 (1), [Accessed 4/16/2018 2:56:21 PM].

Tang, Q., 2012. A GWR- Based Study on Spatial Pattern and Structural Determinants of Shanghai's Housing Price. *Economic Geography,* 32 (2), pp. 52-58. [Accessed 4/11/2018 9:25:08 AM].

Triplett, J. E., 1986. The economic interpretation of hedonic methods. *Survey of Current Business,* 66 (1), pp. 36-40. [Accessed 4/15/2018 2:19:05 AM].

Tse, R. Y. and Love, P. E. 2000. Measuring residential property values in Hong Kong. *Property Management,* 18 (5), pp. 366-374. [Accessed 4/15/2018 8:36:23 AM].

Wang, X., 2015. Spatiotemporal Model Analysis of Urban Housing Price: A Case Study of the Typical Urban of Huangshi City, Hubei Province . *Henan University,* [Accessed 7/26/2018 9:44:10 AM].

Wang, Y. P. and Murie, A. 2000. Social and spatial implications of housing reform in China. *International Journal of Urban and Regional Research,* 24 (2), pp. 397-417. [Accessed 4/2/2018 9:10:17 AM].

Waugh, F. V., 1928. Quality factors influencing vegetable prices. *Journal of Farm Economics,* 10 (2), pp. 185-196. [Accessed 4/15/2018 4:33:42 PM].

Wen, H., 2010. Spatial Variation of Housing Prices in Hangzhou City: Two-Dimensional Analysis Based on Hedonic Prices. *China Land Science,* 24 (2), pp. 51-56. [Accessed 4/11/2018 9:22:53 AM].

Wen, H. and Jia Shenghua 2004. A New Method of Real Estate Evaluation Based on Characteristic Price. *Foreign Economies and Management,* [Accessed 4/16/2018 9:45:13 AM].

Wen, H., Li, X. and Zhang, L. 2012. Impacts of the urban landscape on the housing price: A case study in Hnagzhou. *Geographical Research,* 31 (10), pp. 1806-1814. [Accessed 4/16/2018 9:54:00 AM].

Williams, A. W., 1991. A guide to valuing transport externalities by hedonic means. *Transport Reviews,* 11 (4), pp. 311-324. [Accessed 4/15/2018 9:52:08 AM].

Wu, H., 2016. The Study of Present Developing Situation and Developing Trend of Second-hand Housing Market on First-tier Cities. *Master Thesis, Guangxi University,* [Accessed 4/15/2018 4:27:31 PM].

Xiao, Y., 2017. Hedonic Housing Price Theory Review. Hedonic Housing Price Theory Review. 2017. Urban Morphology and Housing Market. Springer. pp. 11-40. [Accessed 4/16/2018 5:05:31 AM].

Xie, Y., 2013. Regression Analysis. *Social Sciences Academic Press(CHINA),* [Accessed 7/31/2018 3:37:27 AM].

Zhang, J. and Sun, Y. 2016. Analysis of Housing Price Gradient Difference and Its Influencing Factors in Kunming City. *China Market,* (8), pp. 113-115. [Accessed 7/26/2018 9:46:53 AM].

Zhong, M., 2012. Study on the Time and Space Change of Commodity Housing Price in Guangzhou. *Master Thesis, the Guangdong University of Technology,* [Accessed 4/5/2018 7:16:16 AM].

Zhu, C., Tang, Y. and Wu, Q. 2014. A Study on the Influence Factors of the Residential Land Price Based on Hedonic Price Model: Taking Nanjing City as An Example. *Areal Research and Development,* 33 (3), [Accessed 4/15/2018 9:34:36 AM].

Zou, W., Wu, Q. and Qu, F. 2009. Theoretical Analysis and Empirical Research on Influence of Property Tax on the Allocation of Resources . *China Land Science,* 23 (9), pp. 20-24. [Accessed 6/10/2018 3:50:03 AM].

Annex 1: Results of the OLS estimation (Stata 14.2)

```
Linear regression                               Number of obs     =      6,684
                                                F(15, 6668)       =     728.55
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6265
                                                Root MSE          =     .16258
```

lgprice	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
lgsize	.9013767	.0093026	96.90	0.000	.8831406	.9196128
Floors	.0008501	.0002554	3.33	0.001	.0003495	.0013507
Buildingtype	.0442528	.0051341	8.62	0.000	.0341884	.0543172
Buildingorientation	.0643371	.0107025	6.01	0.000	.0433567	.0853175
Houseage	-.0101336	.0009792	-10.35	0.000	-.012053	-.0082141
Plotratio	.0010086	.0019586	0.51	0.607	-.0028309	.004848
Greenratio	.4571584	.0313305	14.59	0.000	.3957405	.5185763
DIST_park	.007134	.003187	2.24	0.025	.0008864	.0133816
DIST_hospital	.0154467	.004568	3.38	0.001	.0064921	.0244014
firstclassprimaryschoold~r	.0292609	.0079696	3.67	0.000	.0136378	.0448839
Numberofbusstopswithin1km	.007227	.0016712	4.32	0.000	.003951	.0105031
DIST_subwaystation	-.0123134	.0013888	-8.87	0.000	-.0150359	-.0095909
DIST_mainroad	-.017843	.0122167	-1.46	0.144	-.0417917	.0061057
DIST_ErqiSquare	.0340122	.0030377	11.20	0.000	.0280574	.039967
DIST_JinshuiRiverPark	-.0506024	.0030251	-16.73	0.000	-.0565326	-.0446722
_cons	9.545981	.0447736	213.21	0.000	9.458211	9.633752

Annex 2: Bandwidth and geographic ranges in the GWR regression (GWR 4.0)

Table 10 Bandwidth and geographic ranges

Coordinate	Min	Max	Range
X-coordinate	113.582	113.746	0.164
Y-coordinate	34.593	34.844	0.151
Bandwidth		940.540	

Annex 3: Coefficients and local R^2 of GWR estimation (Only the first 100 sample points are shown with parts of the whole variables)

ID	Longitude	Latitude	Intercept	lgsize	Floors	Building type	Building orientation	House age	Plot ratio	Green ratio	DIST park	DIST Jinshui Park	residual	std_residual	Local R^2	Cooks D
0	113.674	34.723	9.517	0.872	0.000	0.037	0.114	-0.006	0.011	0.566	0.054	-0.054	-0.020	-0.139	0.755	0.000
1	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.077	0.559	0.727	0.000
2	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	0.135	0.919	0.727	0.000
3	113.677	34.726	9.471	0.871	0.000	0.041	0.116	-0.007	0.009	0.685	0.058	-0.029	0.268	1.855	0.748	0.001
4	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	-0.052	-0.353	0.727	0.000
5	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	0.001	0.008	0.727	0.000
6	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	-0.149	-1.008	0.727	0.000
7	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	0.157	1.063	0.727	0.000
8	113.631	34.733	9.771	0.898	0.000	0.017	0.040	-0.016	0.004	0.107	0.010	-0.036	0.217	1.476	0.727	0.000
9	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.170	1.157	0.727	0.000
10	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.022	-0.150	0.727	0.000
11	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.084	-0.576	0.727	0.000
12	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.146	-1.004	0.727	0.000
13	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.114	-0.783	0.727	0.000
14	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.015	0.102	0.727	0.000
15	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.062	-0.426	0.727	0.000
16	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.013	0.087	0.727	0.000
17	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.112	-0.775	0.727	0.000

18	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.118	-0.804	0.727	0.000
19	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.080	-0.542	0.727	0.000
20	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.074	-0.508	0.727	0.000
21	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.209	1.434	0.727	0.000
22	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.083	-0.567	0.727	0.000
23	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.087	-0.598	0.727	0.000
24	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	-0.080	-0.549	0.727	0.000
25	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.083	0.562	0.727	0.000
26	113.616	34.698	9.863	0.891	-0.001	0.019	0.016	-0.011	0.006	0.140	0.011	-0.098	0.083	0.570	0.727	0.000
27	113.671	34.701	9.614	0.837	0.000	0.025	0.114	-0.002	0.020	0.514	0.063	-0.106	-0.163	-1.095	0.689	0.000
28	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	0.204	1.372	0.679	0.000
29	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.066	-0.443	0.680	0.000
30	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.106	-0.710	0.680	0.000
31	113.636	34.765	9.408	0.927	0.002	0.000	0.046	-0.014	-0.008	0.657	0.010	0.003	-0.246	-1.701	0.676	0.001
32	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.048	-0.325	0.680	0.000
33	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	0.083	0.561	0.679	0.000
34	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.097	-0.653	0.680	0.000
35	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	0.015	0.100	0.680	0.000
36	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.057	-0.387	0.680	0.000
37	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	0.019	0.126	0.679	0.000
38	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	-0.075	-0.509	0.679	0.000
39	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.028	-0.191	0.680	0.000
40	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.061	-0.411	0.680	0.000
41	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	-0.011	-0.074	0.679	0.000

42	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	0.020	0.137	0.680	0.000
43	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	-0.011	-0.078	0.679	0.000
44	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	-0.105	-0.715	0.679	0.000
45	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	-0.128	-0.872	0.679	0.000
46	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.523	0.064	-0.106	0.158	1.066	0.679	0.000
47	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	0.043	0.289	0.680	0.000
48	113.672	34.699	9.615	0.833	0.000	0.027	0.114	-0.001	0.021	0.522	0.064	-0.106	-0.028	-0.187	0.680	0.000
49	113.671	34.720	9.525	0.874	0.000	0.031	0.112	-0.007	0.014	0.486	0.056	-0.063	-0.049	-0.339	0.759	0.000
50	113.672	34.715	9.513	0.857	0.000	0.030	0.113	-0.005	0.018	0.463	0.066	-0.025	0.058	0.393	0.728	0.000
51	113.672	34.715	9.513	0.857	0.000	0.030	0.113	-0.005	0.018	0.463	0.066	-0.025	-0.265	-1.803	0.728	0.000
52	113.672	34.715	9.513	0.857	0.000	0.030	0.113	-0.005	0.018	0.463	0.066	-0.025	-0.232	-1.579	0.728	0.000
53	113.672	34.715	9.513	0.857	0.000	0.030	0.113	-0.005	0.018	0.463	0.066	-0.025	-0.173	-1.180	0.728	0.000
54	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.118	-0.802	0.699	0.000
55	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.123	0.835	0.699	0.000
56	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.181	-1.231	0.699	0.000
57	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.085	-0.579	0.699	0.000
58	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.061	0.418	0.699	0.000
59	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.027	-0.183	0.699	0.000
60	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.433	2.940	0.699	0.001
61	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.156	-1.064	0.699	0.000
62	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.020	-0.139	0.699	0.000
63	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.062	-0.424	0.699	0.000
64	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.078	0.527	0.699	0.000
65	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.067	-0.455	0.699	0.000

66	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.064	0.435	0.699	0.000
67	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	-0.036	-0.248	0.699	0.000
68	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.322	2.189	0.699	0.001
69	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.230	1.563	0.699	0.000
70	113.672	34.706	9.575	0.842	0.000	0.028	0.114	-0.003	0.019	0.496	0.065	-0.068	0.090	0.613	0.699	0.000
71	113.668	34.719	9.537	0.881	0.000	0.026	0.110	-0.008	0.014	0.456	0.054	-0.089	0.100	0.685	0.771	0.000
72	113.668	34.719	9.537	0.881	0.000	0.026	0.110	-0.008	0.014	0.456	0.054	-0.089	0.108	0.741	0.771	0.000
73	113.668	34.719	9.537	0.881	0.000	0.026	0.110	-0.008	0.014	0.456	0.054	-0.089	0.203	1.395	0.771	0.000
74	113.674	34.754	9.424	0.916	0.000	0.051	0.004	-0.010	-0.006	0.696	0.100	-0.017	0.066	0.450	0.757	0.000
75	113.674	34.754	9.424	0.916	0.000	0.051	0.004	-0.010	-0.006	0.696	0.100	-0.017	-0.154	-1.128	0.757	0.001
76	113.651	34.752	9.215	0.970	0.001	0.017	0.076	-0.017	0.000	0.695	-0.009	0.023	0.040	0.276	0.756	0.000
77	113.651	34.751	9.214	0.969	0.001	0.017	0.081	-0.017	0.001	0.694	-0.009	0.022	0.071	0.490	0.760	0.000
78	113.651	34.751	9.214	0.969	0.001	0.017	0.081	-0.017	0.001	0.694	-0.009	0.022	-0.016	-0.113	0.760	0.000
79	113.651	34.751	9.214	0.969	0.001	0.017	0.081	-0.017	0.001	0.694	-0.009	0.022	-0.016	-0.113	0.760	0.000
80	113.651	34.751	9.219	0.969	0.001	0.017	0.079	-0.017	0.001	0.692	-0.010	0.023	0.077	0.533	0.759	0.000
81	113.651	34.751	9.219	0.969	0.001	0.017	0.079	-0.017	0.001	0.692	-0.010	0.023	0.235	1.624	0.759	0.001
82	113.651	34.751	9.214	0.969	0.001	0.017	0.081	-0.017	0.001	0.694	-0.009	0.022	-0.075	-0.520	0.760	0.000
83	113.651	34.752	9.216	0.969	0.001	0.017	0.077	-0.017	0.001	0.694	-0.009	0.023	-0.098	-0.675	0.757	0.000
84	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.078	-0.534	0.776	0.000
85	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	0.166	1.131	0.776	0.000
86	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	0.091	0.622	0.776	0.000
87	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.035	-0.236	0.776	0.000
88	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	0.005	0.037	0.776	0.000
89	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.071	-0.488	0.776	0.000

90	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.116	-0.793	0.776	0.000
91	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.122	-0.836	0.776	0.000
92	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	0.134	0.925	0.776	0.000
93	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	-0.097	-0.661	0.776	0.000
94	113.649	34.745	9.323	0.955	0.001	0.011	0.107	-0.019	0.001	0.619	-0.013	0.024	0.103	0.700	0.776	0.000
95	113.650	34.744	9.318	0.955	0.001	0.012	0.110	-0.019	0.002	0.625	-0.009	0.016	-0.163	-1.114	0.784	0.000
96	113.650	34.744	9.318	0.955	0.001	0.012	0.110	-0.019	0.002	0.625	-0.009	0.016	-0.059	-0.402	0.784	0.000
97	113.650	34.744	9.318	0.955	0.001	0.012	0.110	-0.019	0.002	0.625	-0.009	0.016	-0.095	-0.647	0.784	0.000
98	113.650	34.744	9.318	0.955	0.001	0.012	0.110	-0.019	0.002	0.625	-0.009	0.016	0.074	0.504	0.784	0.000
99	113.650	34.744	9.318	0.955	0.001	0.012	0.110	-0.019	0.002	0.625	-0.009	0.016	-0.048	-0.327	0.784	0.000
100	113.674	34.723	9.517	0.872	0.000	0.037	0.114	-0.006	0.011	0.566	0.054	-0.054	-0.020	-0.139	0.755	0.000

Annex 4: IHS copyright form

In order to allow the IHS Research Committee to select and publish the best UMD theses, participants need to sign and hand in this copy right form to the course bureau together with their final thesis.

Criteria for publishing:

A summary of 300 to 500 words should be included in the thesis.

The number of pages for the thesis is about 60.

The thesis should be edited.

Please be aware of the length restrictions of the thesis. The Research Committee may choose not to publish very long and badly written theses.

By signing this form you are indicating that you are the sole author(s) of the work and that you have the right to transfer copyright to IHS, except for items cited or quoted in your work that are clearly indicated.

I grant IHS, or its successors, all copyrights to the work listed above, so that IHS may publish the work in *The IHS thesis series*, on the IHS web site, in an electronic publication or in any other medium.

IHS is granted the right to approve reprinting.

The author(s) retain the rights to create derivative works and to distribute the work cited above within the institution that employs the author.

Please note that IHS copyrighted material from *The IHS thesis series* may be reproduced, up to ten copies for educational (excluding course packs purchased by students), non-commercial purposes, providing full acknowledgements and a copyright notice appear on all reproductions.

Thank you for your contribution to IHS.

Date : 04-09-2018

Your Name(s) : Chunyue Mao

Your Signature(s) : Chunyue Mao

Please direct this form and all questions regarding this form or IHS copyright policy to:

The Chairman, IHS Research Committee Burg. Oudlaan 50, T-Building 14th floor, 3062 PA Rotterdam, The Netherlands	j.edelenbos@ihs.nl Tel. +31 10 4089851