WILEY | Hindawi

Research Article

House Price Prediction Model of Zhaoqing City Based on Correlation Analysis and Multiple Linear Regression Analysis

Ningyan Chen (iD)

Business School, University of Aberdeen, Aberdeen, UK

Correspondence should be addressed to Ningyan Chen; n.chen.21@abdn.ac.uk

Situated in southern China, Zhaoqing City is a part of Guangdong Province, China. The total administrative area of the city covers 14,891 square kilometers. The data of China's seventh population census in 2020 showed that the permanent resident population in Zhaoqing City reached up to 4,413,594. Meanwhile, Zhaoqing is one of the cities in the Guangdong-Hong Kong-Macao Greater Bay Area. House price analysis and prediction carried out against Zhaoqing City will have directive significance for relevant policies formulated by the local government, residential investment or purchase of consumers, and prediction of house price trend as well as business decisions made by enterprises. By virtue of machine learning and statistical theory, the house price in Zhaoqing City from 2010 to 2020 will be researched, and the house price prediction model of Zhaoqing City will be constructed in this paper with several variables including GDP, proportion of tertiary industry, income of urban residents, fiscal revenue, land price, investment volume in real estate development, permanent resident population, population density, and proportion of urban population in net migration. First of all, the methods of correlation analysis will be utilized, to select variables that are highly correlated with house price data based on correlation coefficients. Then, the model will be constructed for predicting the house price on the basis of multiple linear regression analysis that is conducted with selected variables. Finally, the prediction model will be adjusted gradually based on data with different correlations selected from available data, to realize better imitative effect and more precise predictive effect and select optimum prediction model. By means of the above model, the house prices of Zhaoqing City in 2021 and beyond will be predicted accurately, with preferable fitting effect and prediction effect.

1. Introduction

1.1. China's Real Estate Development. The real estate industry is one of the pillar industries of China's economy and has a significant impact on the national economy. The real estate industry can directly drive the growth of gross domestic product and contribute to China's fiscal revenue. At the same time, the real estate industry also drives the development of upstream and downstream industries such as building materials, machinery manufacturing, finance, and home appliances.

1.1.1. Development of China's Residential Real Estate. The growth of the real estate business can improve people's lives. Land and real estate have always been valued by the Chinese. Every Chinese requires a fixed dwelling, so they are eager to buy properties. Housing is very important to Chinese people. It is not simply a sign of a home but also a requirement for marriage. In this way, some Chinese struggle for a lifetime to buy a home. Real estate development can better meet people's material requirements and improve their quality of life.

Since the People's Republic of China's founding, real estate has been slowly developing. Prior to reform and liberalization, China had a planned economy and welfare housing distribution. A building has several single rooms. Those days, welfare housing had a high population density and a small dwelling area. China's reform and openness era began in 1978. Apartments and villas have replaced tile roofed dwellings and bungalows in China. From 1980 to 1991, China's housing circumstances improved, as did housing quality and facilities. Welfare housing has steadily been

supplanted by commercial housing for personal usage. In 1992 and 1993, China's real estate market overheated. From 1993 to 1997, China lowered property prices. The 1997 economic crisis lowered China's property prices. China's economy and real estate market have both grown rapidly since 1998. China's property values are soaring. At the same time, China's housing market has changed. Environmentally friendly housing has gradually become the preferred choice of Chinese youth, increasing their sense of contentment.

1.1.2. Development of China's Real Estate Investment. Real estate investing includes land and property development, property management, and property buying. To realize the investment advantage, it must mobilize many social resources, including capital, land, material, labor, technology, information, and others. Early stage is from 1978 to 1998, and growth stage is from 1998 to present of China's real estate investment (1999-present).

The Eleventh Central Committee of the Communist Party of China decided to implement reform and opening in 1978, which boosted the development of the real estate business. To begin, China's investigation of the virtuous circle of housing funds modified the national policy of unified building and distribution, enacted public housing sales and rent reform, and opened the bank's real estate credit operations. Second, adopt policies such as public-private collaborative construction, establish housing management cooperatives, rent or sell land with compensation, and increase real estate development. From 1992 to 1998, there was irrational real estate speculation. Then property values soared. Because of this, the state changed its regulations and created the Housing Provident Fund System. This system has enhanced the investment system of real estate investment and established the ground for the growth of policy real estate investment under the market-oriented system. From 1999 to 2015, the arrival of diversified capital chains, new regulatory regulations, and real estate securitization improved real estate investment. Simultaneously, land availability has increased, while demand has soared, resulting in several price-raising behaviors, such as real estate speculation.

Houses are for living, not for speculating, said a December 2016 Chinese economic conference. To accelerate the creation of a long-term mechanism fit for the real estate industry's market law. The Chinese government has used prudent financial, land, fiscal, taxation, investment, and legislative measures to contain the real estate bubble and keep prices stable. The rental sales ratio of property prices is excessively high in China's top tier cities. Due to high housing costs and strong life pressure in first tier cities, many new residents are unable to live a house life. Due to population reduction, house values in second and third tier cities tend to be steady or decline.

2. Brief Introduction of Zhaoqing's Real Estate Market

House prices in China have been increasing in the last 20 years and have a large impact on people's lives. The disparity in house prices between Chinese cities is large. Guangdong Province, located in the south of China, including Guangzhou, Foshan, Zhaoqing, Shenzhen, Dongguan, Huizhou, Zhuhai, Zhongshan, and Jiangmen, is an example. The first-tier cities, represented by Shenzhen and Guangzhou, have very high house prices and are at the forefront of the world. According to CBRE's *Global Living Report 2020*, Shenzhen is in fifth place globally with an average house price of $738 per square foot.

Although Zhaoqing is part of the same Guangdong Province as Shenzhen and Guangzhou, the house price is lower than the provincial average. According to Figure 1, Zhaoqing, located in the central and western part of Guangdong Province, the whole territory is located at $22°47'$ $-24°24'$ N latitude and $111°21'$-$112°52'$ E longitude, with subtropical monsoon climate, and is one of the important node cities in Guangdong-Hong Kong-Macao Greater Bay Area. Zhaoqing City has jurisdiction over Duanzhou District, Dinghu District, Gaoyou District, etc., with economic functional areas such as Zhaoqing High-tech Zone. The total area of the city's administrative region is 14,891 square kilometers [1]. According to the 7th China Census 2020, the resident population of Zhaoqing is 4,413,594. The Guangdong-Hong Kong-Macao Greater Bay Area is located in the south of China and includes Hong Kong, Macau and Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai, Jiangmen, Zhaoqing, Huizhou, and Zhongshan in Guangdong Province, with a total area of $56,000 \, km^2$ and a total population of about 71,159,800.

Figure 2 plotted by the National Bureau of Statistics can visually reflect that the house price in Zhaoqing is lower than the average house price in China and significantly lower than the average house price in Guangdong Province from 2010 to 2020.

The house price trend in Zhaoqing is a microcosm of a part of China's cities with a lower economic level but a greater potential for economic development. According to the *China City Statistical Yearbook* of previous years, since 2010-2020, the house prices in Zhaoqing have risen and then fallen, and total residential sales have been increasing. Figure 3 shows the trend of residential sales and house prices in Zhaoqing from 2010 to 2020.

The residential sales price per square meter in Zhaoqing is influenced by many factors. This research uses correlation analysis, linear regression models, and political, social, and locational factors to establish a prediction model for Zhaoqing house prices based on the relationship between house prices and data on several influencing factors. Establishing the prediction model of Zhaoqing house price is a guide for the government to make policies, consumers to invest or purchase residential properties, enterprises to judge the trend of house price and house price prediction, and upstream and downstream manufacturers to make inventory and business decisions.

The house price prediction model in Zhaoqing is a guide for the Zhaoqing government to introduce policies to control house prices. If house prices are too high and growing too fast, the government can take a series of measures to
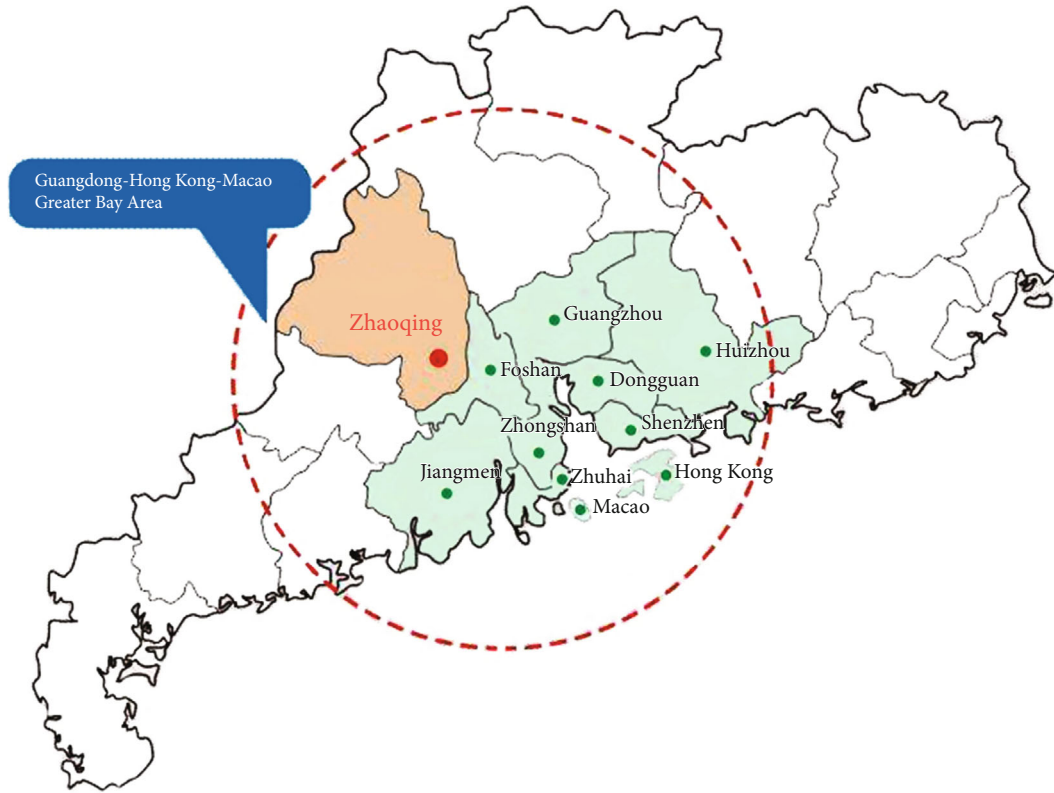
Figure 1: Location of Zhaoqing in the Guangdong-Hong Kong-Macao Greater Bay Area.
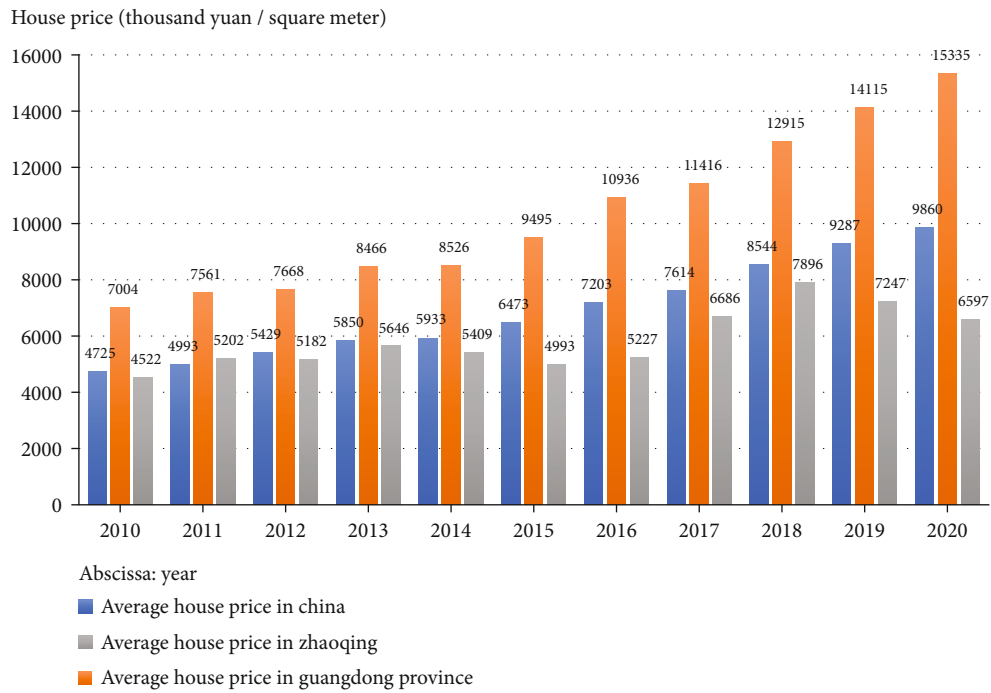

Figure 2: Comparison of average house prices in China, Guangdong Province, and Zhaoqing City.

control house prices; if house prices continue to fall, it is likely to have a negative impact on the local economy, and the government can make preparations for a reduction in real estate tax revenue. At the same time, the house price prediction model can make the house price prediction model can make the government aware of the gap between

House price of zhaoqing (thousand yuan / square meter)   Residential sales of zhaoqing (100 million yuan)

Abscissa: year

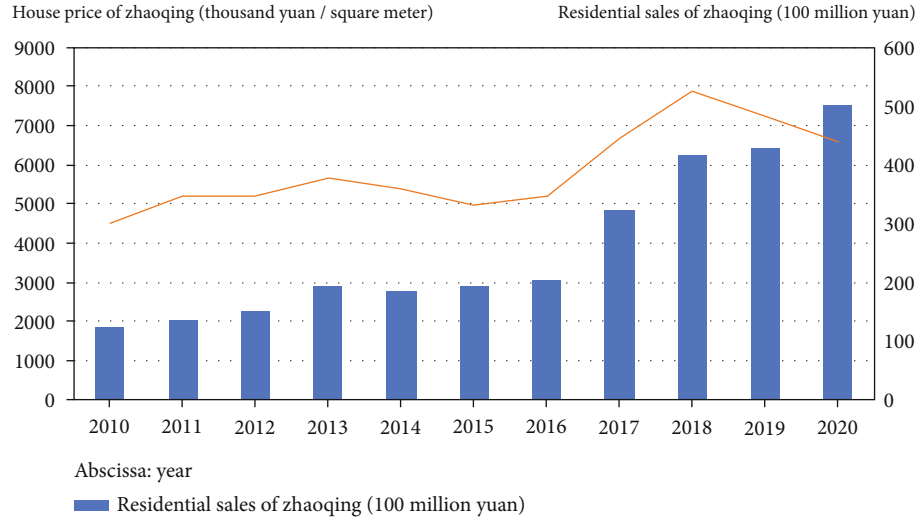■ Residential sales of zhaoqing (100 million yuan)

FIGURE 3: Residential sales and house prices of Zhaoqing from 2010 to 2020.

house prices and residents' income and vigorously develop the economy to make up for this gap and improve residents' quality of life. Stable house prices can make Zhaoqing's economy run more smoothly.

In addition, home price prediction models guide consumers in purchasing or investing in homes. The house price prediction model has implications for consumers, who can determine whether house prices will rise or fall in the future and the magnitude of the change in house prices, etc., and calculate the rate of return on investment in real estate. Consumers can also determine whether it is the right time to buy a home as an immediate need and whether the market is at a lower price at this time, and residents can decide whether to buy a local home in that year based on this model.

From the industrial perspective, for companies, advance judgment of house price trends and house price forecasts can bring some value, and upstream and downstream manufacturers can prepare inventory and make business decisions. Companies can judge rental costs based on the house price forecast model, and investment companies can judge future earnings based on this model. Industries that are highly correlated with real estate prices need to determine real estate prices and rents in advance, such as manufacturing and retail, and the house price prediction model enables companies to make better decisions.

3. Literature Review

There are many classic theories affecting house prices. This paper introduces location theory, land rent and land price theory, and land return decline theory.

The neoclassical economist Alfred Marshall (2009) put forward the location theory to analyze the factors such as economy, society, politics, climate, and location, which reflects the correlation degree of the mutual operation relationship of location subjects in spatial location. Comprehensive location factors have a certain impact on house prices, and location theory can better analyze house prices [2]. Petty put forward the theory of land rent and land price. Then,

there is the emergence of Western neoclassical urban land rent theory, which analyzes the spatial distribution characteristics of urban land prices. This theory can better analyze and compare the house prices in different regions of the city [1]. Malthus summarized the theory of diminishing land returns and analyzed the view that the annual output increased according to the proportion of farming progress must gradually and invariably decrease compared with the previous average increase. This theory plays a guiding role in real estate investment and development [3].

In recent years, with the development of statistics and machine learning, many researchers have carried out quantitative analysis on the variables that affect housing prices, which has realized the prediction of housing prices. Some researchers have used the same multivariable linear regression model as this paper to predict housing prices in different cities, which has achieved good results. Also, some researchers have used other models using machine learning and data mining algorithms, including ANN model, SVM, ARIMA model, Neural Network, C&R Tree, Multilayer Perception model, GM (1,1) model, and Hedonic model to predict the housing price of some cities in the Guangdong-Hong Kong-Macao Greater Bay Area of China, which also has achieved great results.

In recent years, scholars have used linear regression models to predict housing prices in different cities around the world. Ghosalkar and Dhage selected the data of three factors affecting house prices, including physical conditions, concepts, and locations. Without any expectation of market price and cost increment, they used linear regression to predict house prices in Mumbai, India [4]. Kaushal et al. compared the accuracy of some machine learning models, such as multiple linear regression, Lasso, ridge, and decision tree regression models, to find out which one has the best performance. The results showed that the multiple linear regression model selected in this paper has the highest accuracy compared with other machine learning models [5]. Amri and Tularam established multiple regression, linear regression, and nonlinear models to compare their effects on

predicting house prices in Bathurst, Australia. The results show that the improved linear MR performed almost as well as the nonlinear NN. In most situations, linear approaches appear to be equally as exact as the more time-consuming nonlinear methods for accounting for variances and variation [6].

In addition, some scholars have used other models of machine learning to predict housing prices in cities in the Guangdong-Hong Kong-Macao Greater Bay Area, such as Hong Kong, Macau, Shenzhen, and Guangzhou. Abidoye et al. collected variables affecting house prices in Hong Kong, such as interest rate, unemployment rate, and family size, and fitted them with ARIMA, ANN, and SVM models to generate forecasts of real estate prices [7]. Fong and Wah gathered multiattribute datasets from the Macao SAR Government's Statistics and Census Service and used several data mining methods and algorithms such as SVM, Neural Network, C&R Tree, Weka, SPSS, and Multilayer Perception model to forecast Macao home prices [8]. Hongbin constructed and compared the prediction accuracy of ARIMA model and GM (1,1) model on house prices in Shenzhen and judged which model has better effect in predicting the change trend of house prices in the future [9]. Bilei used the Hedonic model to analyze the relationship between honorary title and house price in the Guangzhou community, and the model test results are good [10].

Previous scholars used linear regression and other machine learning models to analyze and predict housing prices in many cities. However, there is currently no research on the housing price prediction on Zhaoqing City in the Guangdong-Hong Kong-Macao Greater Bay Area of China. Based on correlation analysis and multiple linear regression analysis, this study provides a set of methodology as well as a model for predicting housing prices in Zhaoqing. This model is of guiding significance for the Zhaoqing government to formulate policies, consumers to invest or purchase houses, and enterprises to judge the trends of housing price.

4. Research Hypothesis and Analysis of Factors Affecting the House Price of Zhaoqing

4.1. Research Hypothesis. We hypothesize that collecting the data of multiple factors affecting the house price in Zhaoqing can predict the house price. Moreover, the house price of Zhaoqing can be expressed as a formula by multiple influencing factors.

The factors affecting the house price in Zhaoqing are divided into quantifiable and nonquantifiable factors. Quantifiable factors, such as GDP and population size, can be converted into data that visually reflect the variable and form a formula to predict the house price. The nonquantifiable factors, such as policy factors, which can also be used to predict the house price in Zhaoqing, need to be considered outside the model.

4.2. Quantifiable Factors. The following factors that affect house prices can be found in the relevant data, and the researcher can quantify the factor. Based on the data below, a model is developed using correlation and multiple regression methods that can be used to predict future house prices in Zhaoqing City.

4.2.1. Economic Factors

(1) Economic Development Level (GDP) and Fiscal Revenue. GDP is the final result of production activities of all resident units in a country or region in a certain period of time [11]. GDP of Zhaoqing, i.e., the value of production activities in Zhaoqing in a certain period of time, can be used to measure the level of economic development of Zhaoqing. Fiscal revenue, the sum of all funds raised by the government to carry out its functions, implement public policies and provide public goods and services as needed [12]. Zhaoqing's GDP influences Zhaoqing real estate prices. The higher Zhaoqing's GDP, the more developed the local economy is, and the more valuable Zhaoqing real estate is. The price is determined by the value, and the more expensive the house in Zhaoqing, the higher the price. Higher fiscal revenues indicate higher government revenues through taxes and other means, representing a more dynamic local economy with more business activity.

(2) Industrial Structure. Industrial structure is a concept proposed in development economics. Industrial structure, also called industrial system, is the main component of the socioeconomic system, including agriculture, industry, and services, i.e., primary, secondary, and tertiary industries [13]. Figure 4 is a graph of the ratio of the output value of primary, secondary, and tertiary industries in GDP of Zhaoqing from 2010 to 2020, based on the data of Statistics Bureau of Guangdong Province.

It can be seen from the above graph that the output value ratio of primary industry in Zhaoqing is the lowest from 2010 to 2020. In 2010-2016, the share of output value of tertiary industry in Zhaoqing is lower than the share of output value of secondary industry. However, in 2017-2020, the share of output value of tertiary industry in Zhaoqing is higher than the share of output value of secondary industry. These data indicate that the proportion of tertiary industry output value in Zhaoqing has increased in recent years. The higher ratio of tertiary industry means the higher quality of economic development and the more attractive to foreign workers. High-end service industry can attract talents, and the concentration of population makes the demand for housing higher. At the same time, the developed tertiary industry provides a certain number of supporting facilities for housing, such as restaurants, entertainment and recreation, and sports. The secondary industry, on the other hand, has lower wage levels than the tertiary industry, fewer workers, and difficulty in population clustering, which leads to a lower demand for housing. Therefore, the proportion of both secondary and tertiary industries in Zhaoqing is likely to be a factor affecting the house price in Zhaoqing.

(3) Income of Urban Residents and Level of Employment of Residents. The income of urban residents is the sum of value created by workers in a city over a certain period of time
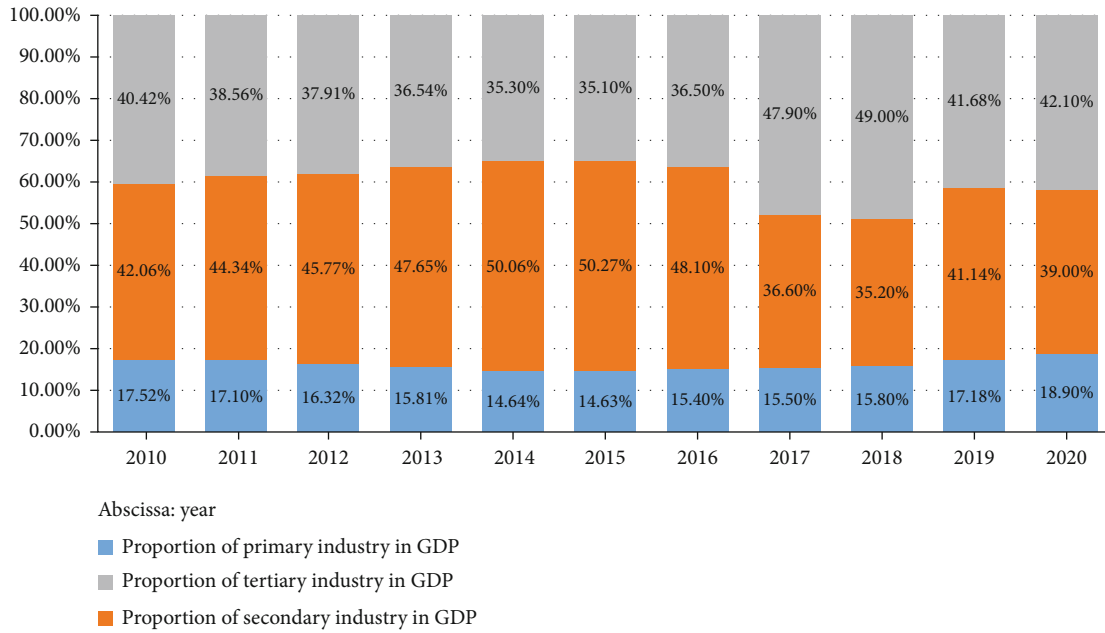
FIGURE 4: Proportion of Zhaoqing's primary, secondary, and tertiary industries in GDP.

[14]. The employment level of residents is the number of employed population. The employed population is the population aged 16 years and above who are engaged in certain social labor or business activities and receive labor remuneration or business income [15]. The higher the disposable income of urban residents, the better the local economy, the stronger the willingness of residents to purchase residential housing, and the higher the demand for residential housing. The higher the employment level of residents, the more the local working population, and the stronger the ability of the working population to buy residential houses. Therefore, the income of urban residents and the employment level of residents in Zhaoqing may have some influence on the house price in Zhaoqing.

(4) *Land Price and Investment in Real Estate Development.* The land price is the price of a piece of land or a tract of land at a certain point in time in a certain state of rights [16]. The price of land is the cost of building a home. The higher the cost of constructing a residence for a developer, the higher the selling price of the house may be accordingly. In addition, the planning of residential land in Zhaoqing City will also affect the price of housing.

Real estate development investment is the amount of investment completed in a certain period of time, and real estate development companies and commercial housing construction companies matter real estate development or business activities of the unit of unified development [17]. If the amount of investment in real estate development in Zhaoqing is increased, it may be that developers invest more costs in individual properties, or developers may build more properties in Zhaoqing, which may have an impact on the price of housing in Zhaoqing.

4.2.2. Social Factors

(1) *Population (Resident Population, Population Density, and Net Migration of Population).* The resident population is defined as the population that is always at home or resides at home for more than 6 months throughout the year and also includes mobile people who live in the city where they are located. Population density is the number of people per unit of land area. In addition, it is an important indicator of the distribution of population in a country or region. The higher the population density, the more concentrated the population is. The more concentrated the population is, the more productive the place is, the higher the economic level is, and the higher the house price is likely to be. The population migration is the absolute amount of population moving in and out of a certain period and a certain area [18]. Net population migration is to some extent indicative of the local economic situation and the income level of local residents, and this factor may have some impact on house prices.

(2) *Urbanization (Proportion of Urban Population).* The degree of urbanization is measured by the proportion of urban population. Urbanization refers to the historical process of gradual transformation of a country or region's society from a traditional rural-type society, which is mainly agricultural, to a modern urban-type society, which is mainly nonagricultural industries such as industry and services, with the development of social productivity, progress in science and technology, and the adjustment of industrial structure [19]. The proportion of urban residents reflects the stage of development of local urbanization. The greater the degree of urbanization, the more developed the economy, and consequently, the more expensive the home.

TABLE 1: Data on factors affecting house prices in 2010-2020 of Zhaoqing.

(a)

Year	Y (house price in Zhaoqing)	X_1 (GDP)	X_2 (proportion of secondary industry)	X_3 (proportion of tertiary industry)	X_4 (income of urban residents)	X_5 (resident employed persons)	X_6 (fiscal revenue)
2010	4522	1085.87	42.06	40.42	16832.37	213.05	76.8
2011	5202	1323.3	44.34	38.56	19039.65	215.13	92.23
2012	5182	1473.84	45.77	37.91	21754.4	215.55	103.81
2013	5646	1660.07	47.65	36.54	23929.84	216.22	120.77
2014	5409	1845.06	50.06	35.3	21725.8	217.79	139.1
2015	4993	1970.01	50.27	35.1	18991.4	218.44	143.36
2016	5227	2084.02	48.1	36.5	25907.21	220.31	91.7
2017	6686	2110.01	36.6	47.9	28276.1	221.31	94.85
2018	7896	2201.8	35.2	49	30679.6	225.3	106.04
2019	7247	2248.8	41.14	41.68	33259.8	231.64	114.21
2020	6597	2311.65	39.81	42.7	35470.16	235.08	121.99

(b)

Year	X_7 (land price)	X_8 (investment in real estate development)	X_9 (resident population)	X_{10} (population density)	X_{11} (net migration of population within the province)	X_{12} (net migration of population outside the province)	X_{13} (proportion of urban population)
2010	385.64	91.14	392.22	273.16	8528	2204	42.39
2011	1306.22	143.11	395.14	276.06	2497	2653	42.45
2012	947.15	145.45	398.23	287.15	-4497	858	42.62
2013	935.16	171.48	402.21	288.64	-17228	-73	43.82
2014	466.46	188.89	403.58	291.27	-5731	-1026	44.01
2015	580.64	159.77	405.96	294.14	-5521	-1405	45.16
2016	849.52	145.2	408.46	1369	-8250	-728	46.08
2017	2545	208.04	411.54	1707	-9175	3340	46.78
2018	1763	348.03	415.17	1974.02	-13086	2944	47.76
2019	2095	456.79	418.71	2045.84	-15580	2664	48.63
2020	2081	586.66	421.71	2343.67	-18467.05	1697.36	49.4

(Data source: National Bureau of Statistics, Statistics Bureau of Guangdong Province, Zhaoqing Municipal Government) (unit: GDP (billion yuan), proportion of secondary industry (billion yuan, %), proportion of tertiary industry (billion yuan, %), income of urban residents (yuan), number of employed residents (10,000 persons), fiscal revenue (billion yuan), land price (yuan/m^2), investment in real estate development (yuan/m^2), resident population (10,000 person), population density (person/km^2), net migration of population within the province (person), net migration from outside the province (person), and urban population ratio (%)).

Table 1 is a list of factors that can be quantified to influence house prices in Zhaoqing. Since data prior to 2010 is more difficult to obtain from public sources, only data on factors affecting house prices in 2010-2020 are used.

4.3. Nonquantifiable Factors. In addition to quantifiable economic and demographic factors, there are other nonquantifiable factors that affect the house prices in Zhaoqing, such as political factors like policies and planning and social factors like transportation and education and location, among others. Since these factors are not easily quantifiable, such factors are not analyzed in depth in the model part of the research, but they can be considered as supplementary elements in the practical operation of house price analysis and forecasting.

4.3.1. Political Factors

(1) Policies of Guangdong-Hong Kong-Macao Greater Bay Area. Benchmarked against the San Francisco Bay Area and New York Bay Area in the United States, the Guangdong-Hong Kong-Macao Greater Bay Area is an important strategy for China to promote in-depth cooperation between the Mainland, Hong Kong, and Macao and is one of the most open and economically vibrant regions in China. Relying on the Guangdong-Hong Kong-Macao Greater Bay Area, Zhaoqing City has certain development potential and space.

Zhaoqing makes full use of the Guangdong-Hong Kong-Macao Bay Area's regional and economic advantages. As can be seen, Zhaoqing's economy will continue to grow in the

future. Zhaoqing will continue to attract more employed people and investors because of Guangdong-Hong Kong-Macao Bay Area policies, and Zhaoqing's house prices will be significantly influenced in the future by Guangdong-Hong Kong-Macao Bay Area.

(2) Housing Policy and Urban Planning. The residential purchase policies in the Guangdong-Hong Kong-Macao Greater Bay Area differ from each other. Since Hong Kong and Macau use different real estate market policies than Mainland China, only cities in the Pearl River Delta are selected for residential purchase policy comparison in this research. The Pearl River Delta in Guangdong Province is one of the strongest city clusters in China, consisting of nine cities with high economic levels in Guangdong Province, Guangzhou, Foshan, Zhaoqing, Shenzhen, Dongguan, Huizhou, Zhuhai, Zhongshan, and Jiangmen. Table 2 provides a visual comparison of the residential purchase policies of Zhaoqing City and other cities in the Pearl River Delta.

Compared to other cities in Guangdong Province, Zhaoqing is the only city that does not restrict the purchase of residential units for both local and foreign household members. This is due to the lower economic level of Zhaoqing compared to other cities in the Guangdong-Hong Kong-Macao Greater Bay Area and the lower demand for purchasing residential properties. But at the same time, Zhaoqing's real estate has a huge potential for development.

In addition, Zhaoqing City has reasonably controlled the scale of residential land in recent years. According to the *Urban Master Plan of Zhaoqing New Area, Guangdong (2012-2030)* and the *Urban Master Plan of Zhaoqing (2015-2030)*, the planned urban construction land in the new district is 6,046.1 hectares, with 1,740.9 hectares of residential land, accounting for 28.8% of the urban construction land. This provides space for property developers to build new residences.

4.3.2. Social Factors

(1) Culture and Tourism. Zhaoqing has a relatively deep historical and cultural heritage and rich tourism resources. Zhaoqing Star Lake Scenic Area is a National AAAAA level tourist attraction in China [20]. Based on the rich tourism resources, Zhaoqing can vigorously develop tourism industry, pension industry, etc., which will bring more economic benefits.

(2) Transportation. Zhaoqing is the only regional center city in the Greater Bay Area bordering the Great Southwest of China and is positioned as a hub gateway connecting the Greater Bay Area to the Great Southwest. In recent years, Zhaoqing has been encrypting its high-speed rail and urban rail network, and it has become a reality to travel from Zhaoqing to Foshan in 20 minutes, Guangzhou in 40 minutes, and Shenzhen in 1 hour and Hong Kong in 1 hour and 20 minutes [21]. This enhances Zhaoqing's overall transportation capacity, accelerates the agglomeration of

people, logistics, capital, and information, and promotes economic development.

(3) Public Service Facilities and Education. Zhaoqing plans to improve public service facilities such as land for administrative offices, cultural facilities, education and research, sports, medical and health care, and social welfare facilities. In recent years, Zhaoqing's public service facilities are being gradually improved. In addition, Zhaoqing has a stable education workforce and an overall increase in expenditure costs for education, which basically meets the needs of basic education.

4.3.3. Locational Factors

(1) Radiation Influence of Surrounding Cities. Zhaoqing relies on the Guangdong-Hong Kong-Macao Greater Bay Area, with Guangzhou City and Foshan City of relatively high economic level nearby, forming a pattern of Guangzhou, Foshan, and Zhaoqing as one city. Guangzhou, as the national central city, has a strong regional radiation carrying power. The central city of Foshan and the central city of Zhaoqing, as subcenters, carry the comprehensive service function of leading the comprehensive and balanced growth of the economic circle. In recent years, Zhaoqing has been developing its economy by taking advantage of its resources and using the radiation drive of Guangzhou and Foshan.

(2) Climate. Zhaoqing is located on the south side of the Tropic of Cancer and is a humid climate region influenced by the southern subtropical monsoon. With abundant rainfall, abundant sunshine and mild climate, the annual average temperature of 22°C, Zhaoqing has a good climate environment for people to work and live [22]. Due to the local cold, some northerners in China purchase homes in the south, which may have an effect on Zhaoqing's housing prices.

5. Research Method

The Pearson coefficient correlation analysis and multiple linear regression analysis were chosen for this research. Firstly, the variables with high correlation with the house prices in Zhaoqing City in 2010-2018 were found by correlation analysis and then analyzed based on the multiple linear regression model to obtain the value of the goodness of fit R^2. Then, according to the formula of the multiple linear regression model, the difference between the predicted and actual house prices for the two years of 2019 and 2020 is obtained $|D|$. Finally, the value of the goodness-of-fit R^2 and the difference between the predicted and actual house prices of house prices $|D|$ are combined to observe the prediction effect.

This research first selected all variables with correlation coefficients $|R|$ greater than 0.8 for the years 2010-2018 to build multiple regression model A. In order to enrich and improve the research, I debugged the model. In the process of adjusting the model, this research selected variables with correlation coefficients $|R|$ greater than 0.79 and 0.78 from 2010 to 2018 to build multiple regression models B and C.

TABLE 2: Residential purchase policies of the Pearl River Delta.

City	Number of restricted residences of local registered population	Number of restricted residences of nonlocal registered population	Whether to restrict the purchase of second-hand house
Zhaoqing	No restriction	No restriction	No
Guangzhou	Single person is restricted to one suite, a family is restricted to two suites	Pay social security for 5 years or more, restricted to one suite	Yes
Shenzhen	Single person is restricted to one suite, a family is restricted to two suites	Pay social security for 5 years or more, restricted to 1 set	Yes
Foshan	Restricted to two suites	Pay social security for 1 years or more, restricted to one suite	No
Dongguan	Restricted to two suites	Restricted to 2 suites (pay social security for 1 year or more, the first suite can be bought; pay social security for 2 years or more, the second suite can be bought.)	No
Huizhou	Key areas (Huiyang, Daya Bay) are restricted to one suite	Restricted to one suite	No
Zhuhai	Restricted to three suites	Restricted to one suite	Yes
Zhongshan	Restricted to three suites	Pay social security for 2 years or more, restricted to two suites	No
Jiangmen	Restricted to three suites	Pay social security for half a year or more, restricted to one suite, pay social security for one year or more, restricted to two suites	No

I selected variables with correlation coefficients $|R|$ greater than 0.8 and 0.77 from 2010 to 2019 to build multiple regression models D and E. Finally, I selected the most suitable model for predicting Zhaoqing house prices by comparing R^2 and $|D|$.

5.1. Model Introduction

5.1.1. Correlation Coefficient. This research uses Pearson correlation coefficient to investigate the relationship between multiple variables and house prices in Zhaoqing.

The overall correlation coefficient ρ is defined as the ratio of the covariance between two variables X and Y, and the product of the standard deviation of the two is as follows:

$$\rho_{(X,Y)} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \quad (1)$$

Estimating the covariance and standard deviation of the sample gives the sample Pearson correlation coefficient, often expressed by r:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (2)$$

r can also be estimated from the standard score means of the (X_i, Y_i) sample points to obtain an expression equivalent to the above equation:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma X}\right)\left(\frac{Y_i - \bar{Y}}{\sigma Y}\right). \quad (3)$$

In the equation, $X_i - \bar{X}/\sigma X$, $\bar{X}$ and σX are the standard score of sample X_i, sample mean, and sample standard deviation, and n is the number of samples [23].

5.1.2. Multiple Linear Regression

(1) Model Assumption. The linear assumption means that the mean of the distribution obeyed by the dependent variable can be expressed using a linear combination of the independent variable and its associated term.

The assumption of normality is the most fundamental assumption that distinguishes linear models from other models. It means that the dependent variables corresponding to each sample obey normal distribution. Combined with the linearity assumption, the linear model can be written in the following form:

$$y_i \sim N(\hat{y}_i, \sigma^2). \quad (4)$$

Because the residuals of the model $\varepsilon_i = y_i - \hat{y}_i$, so there is $\varepsilon_i \sim N(0, \sigma^2)$. The normality assumption actually requires that the residuals of the model corresponding to each sample obey the same normal distribution. If y does not obey a normal distribution, it is not a linear model.

(2) Establishment of Model. Multiple linear regression is a model that predicts quantitative Y based on multiple predictor variables X. It assumes an approximate linear relationship between X and Y. We can write this linear relationship as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon. \quad (5)$$

In equation (5), β_0 expresses the intercept term in the

linear model, and $\beta_1, \beta_2, \beta_3 \cdots \beta_n$ express the slope term. We use the estimated values of the model coefficients β_0 and $\beta_1, \beta_2, \beta_3 \cdots \beta_n$, and it is possible to use the specific variable X to predict the future quantitative Y.

$$y^\wedge = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \cdots \widehat{\beta}_p x_p + \varepsilon. \tag{6}$$

In the equation, $y^\wedge$ expresses the value of y based on the prediction of $X = x$.

(3) Model Solution. In the multiple linear regression equation, a matrix is usually used to solve the equation. The weight of the independent variable X with respect to the dependent variable Y, in linear algebra, can be written in the following form:

$$[\beta_0 \beta_1 \beta_2 \cdots \beta_n] \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \beta^T x = h_\beta(x). \tag{7}$$

The results from linear regression should satisfy the following relationship with the actual results:

$$y^{(i)} = \beta^T x^i + \varepsilon^{(i)}. \tag{8}$$

Then, the probability distribution density function of the error can be written as follows:

$$p\left(\varepsilon^{(i)}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^{(i)}}{\sqrt{2}}\right)^2. \tag{9}$$

In addition, the maximum likelihood function is as follows:

$$L(\beta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\frac{y^{(i)} - \beta^T x^{(i)}}{\sqrt{2}}\right)^2\right). \tag{10}$$

Express the square of the mean error of the model

$$J(\beta) = \frac{1}{2} \sum_{i=1}^{m} \left(y^{(i)} - \beta^T x^{(i)}\right)^2. \tag{11}$$

It can be expressed in matrix as follows:

$$J(\beta) = \frac{1}{2} (x\beta - y)^T (x\beta - y). \tag{12}$$

Taking the partial derivative of equation (12), x is regarded as a column vector, and let $\delta J(\beta)/\delta \beta = 0$, the solution is as follows:

$$\beta = \left(x^T x\right)^{-1} x^T y. \tag{13}$$

The best β was solved above using matrix operations, and the best β will be solved by gradient descent method next, and the results will be analyzed [24].

(4) Model Evaluation. The research performed multiple regression analysis on multiple independent and dependent variables, and the data of multiple regression models with multiple indicators could be obtained. The indicators of the multiple regression model include Multiple R, R Square, Adjusted R Square, standard error, Observed Value, df, SS, MS, Significance F, Intercept Coefficients, and X Variable Coefficients. Below is my interpretation of the multiple linear regression model for each metric is interpreted below.

First, I would like to introduce the metrics of the regression statistics section; Multiple R refers to the correlation coefficient R, whose value is between -1 and 1. The closer the value of Multiple R is to 1, the higher the positive correlation, and vice versa, the higher the negative correlation. Adjusted R Square refers to the corrected coefficient of determination. When researchers compare two regression equations with different numbers of independent variables, they must also consider the effect of the number of independent variables included in the equation. The standard error is equal to the square root of the residual SS/residual df. This, like the coefficient of determination, describes how well the regression model fits the actual data, and it represents the distance between the actual values and the regression line. The observations, on the other hand, refer to how many sets of independent variables there are.

Then, I want to introduce the metrics of the analysis of variance (ANOVA). Regression analysis df refers to the degree of freedom of the regression analysis model, the number of independent or free to vary in the sample when the overall is estimated from the sample. The data degrees of freedom are equal to the number of sample groups minus 1. The degrees of freedom of the regression analysis model are 1, that is, this regression model has 1 parameter, and the residual degrees of freedom are equal to the total degrees of freedom minus the degrees of freedom of the regression analysis model. SS is the sum of error squares, and MS is the mean squared deviation. Significance F refers to the p value of the F-test. Subtracting this value from 1 is the confidence level. When the p value is less than 0.05, the confidence level of the regression model is greater than 95%.

The last metric I would like to introduce is the regression parameter table. t Stat refers to the regression coefficient or standard error. t-test refers to the correlation of a variable with the dependent variable Y. Intercept Coefficients refers to the regression value of the intercept. X Variable Coefficients refers to the regression value of the slope. p value is the p value corresponding to the t-test [25].

The main metrics used in this paper are R Square, Intercept Coefficients, and X Variable Coefficients. After obtaining the individual data of the multiple regression model,

we use R Square to observe the magnitude of the model fit and thus select the best prediction model. Then, the formulas of this model are given using Intercept Coefficients and X Variable Coefficients, which are used to predict the future house prices in Zhaoqing City.

(5) Application Area of Model. Linear regression models have important applications in the fields of biology, medicine, economics, management, agriculture, industry, engineering, and technology, in addition to the field of economics.

5.2. Data Analysis

5.2.1. Correlation Analysis. The relationship between house prices (Y variable) and influencing factors (X variable) in Zhaoqing from 2010 to 2018 was calculated using Pearson correlation coefficient as Table 3.

Select from |Pearson correlation coefficient | variable > 0.8, shown in Table 4.

5.2.2. Multiple Regression Analysis. Multiple regression analysis is carried out between the selected five variables and house price in Zhaoqing as Table 5.

Obtain $R^2_A = 0.99506$.

$$Y_{2019} = -6407.34 + 77.48 * X_1 + 0.10 * X_2 \\ + 26.99 * X_3 + 6.89 * X_4 + (-0.44) * X_5. \tag{14}$$

Substituting the data for 2019 to test the model,

$$Y_{2019} = -6407.34 + 77.48 * 41.68 + 0.10 * 33259.8 \\ + 26.99 * 231.64 + 6.89 * 456.79 \\ + (-0.44) * 2045.84 = 8647.08. \tag{15}$$

Substituting the data for 2020 to test the model,

$$Y_{2020} = -6407.34 + 77.48 * 42.7 + 0.10 * 35470.16 \\ + 26.99 * 235.08 + 6.89 * 586.66 \\ + (-0.44) * 2343.67 = 9803.75. \tag{16}$$

Average prediction difference

$$D_A = \frac{(D_{2019} + D_{2020})}{2} \\ = \frac{\left| (Y_{2019\,actual} - Y_{2019\,predicted}) + (Y_{2020\,actual} - Y_{2020\,predicted}) \right|}{2} \\ = |(7247 - 8647.08) + (6597 - 9803.75)|/2 = 2303.46. \tag{17}$$

5.3. Model Debugging Records. In order to select a relatively good model to predict house prices, different models need to be built by selecting different variables and using data from different years. Then, the model for predicting future house prices in Zhaoqing City is selected by comparing R^2 and the difference between actual and forecast.

TABLE 3: Relationship between house prices and influencing factors from 2010 to 2018.

Pearson correlation	House price in Zhaoqing
GDP	0.670877
Proportion of secondary industry	-0.73752
Proportion of tertiary industry	0.819877
Income of urban residents	0.891052
Employed persons of residents	0.866204
Fiscal revenue	0.012281
Land price	0.763675
Investment in real estate development	0.956127
Resident population	0.798563
Population density	0.824473
Net migration of population within the province	-0.62069
Net migration of population outside the province	0.495307
Proportion of urban population	0.787166

5.3.1. Adjust the Model with Data of 2010-2018. In the process of testing the model for 2010-2018, the variables with correlation coefficients $|R|$ greater than 0.8 were first selected to build the model separately. To debug the model, we then selected the influential variables with $|R|$ greater than 0.79 and 0.78 to build two other regression models, respectively, and then tested them for comparison based on the data for the two years 2019 and 2020.

(1) Select |Pearson Correlation Coefficient | Variable > 0.79. In the data from 2010 to 2018, the variables selected with | Pearson correlation coefficient | >0.79 are as Table 6.

Multiple regression analysis is carried out between the selected six variables and house price in Zhaoqing as Table 7.

Obtain $R^2_B = 0.99506$.

$$Y = -6447.56 + 77.52 * X_1 + 0.10 * X_2 + 27.32 * X_3 \\ + 6.88 * X_4 + (-0.44) * X_5 + (-0.08) * X_6. \tag{18}$$

Substituting the data for 2019 to test the model,

$$Y_{2019} = -6447.56 + 77.52 * 41.68 + 0.10 * 33259.8 \\ + 27.32 * 231.64 + 6.88 * 456.79 \\ + (-0.44) * 2045.84 + (-0.08) * 418.71 = 8646.91. \tag{19}$$

Substituting the data for 2020 to test the model,

$$Y_{2020} = (-6447.56) + 77.5207 * 42.7 + 0.095203 * 35470.16 \\ + 27.32024 * 235.08 + 6.88196 * 586.66 \\ + (-0.43857) * 2343.67 + (-0.08477) * 421.71 \\ = 9635.64. \tag{20}$$

Table 4: |Pearson correlation coefficient | variable > 0.8 between house prices and influencing factors from 2010 to 2018.

Pearson correlation	House price in Zhaoqing
Proportion of tertiary industry	0.819877
Income of urban population	0.891052
Employed persons of residents	0.866204
Investment in real estate development	0.956127
Population density	0.824473

Table 5: Multiple regression analysis of |Pearson correlation coefficient | variable > 0.8 between house prices and influencing factors from 2010 to 2018.

Regression statistics				
Multiple R			0.997527	
R Square			0.99506	
Adjusted R Square			0.986827	
Standard error			118.1515	
Observed Value			9	
Variance analysis				
	df	SS	MS	Significance F
Regression analysis	5	8435879	1687176	0.001174
Residual error	3	41879.33	13959.78	
Total	8	8477758		
Regression parameter				
	Coefficients	Standard error	t Stat	p value
Intercept	-6407.34	15908.11	-0.40277	0.7141
X variable 1	77.47594	27.49129	2.8182	0.066838
X variable 2	0.095128	0.028803	3.302685	0.045641
X variable 3	26.98588	70.52192	0.382659	0.727474
X variable 4	6.88899	2.611939	2.6375	0.077823
X variable 5	-0.43757	0.401007	-1.09118	0.35499

Table 6: |Pearson correlation coefficient | variable > 0.79 between house prices and influencing factors from 2010 to 2018.

Pearson correlation coefficient	House price in Zhaoqing
Proportion of tertiary industry	0.819877
Income of urban residents	0.891052
Employed persons of residents	0.866204
Investment in real estate development	0.956127
Population density	0.824473
Resident population	0.798563

Table 7: Multiple regression analysis of |Pearson correlation coefficient | variable > 0.79 between house prices and influencing factors from 2010 to2018.

(a)

Regression statistics	
Multiple R	0.997527
R Square	0.99506
Adjusted R Square	0.98024
Standard error	144.7054
Observed Value	9

(b)

Variance analysis	df	SS	MS	Significance F
Regression analysis	6	8435879	1405980	0.014747
Residual error	2	41879.29	20939.65	
Total	8	8477758		

(c)

Regression parameter	Coefficients	Standard error	t Stat	p value
Intercept	-6447.56	37221.94	-0.17322	0.878424
X variable 1	77.5207	48.77848	1.58924	0.252954
X variable 2	0.095203	0.068862	1.38253	0.300949
X variable 3	27.32024	277.4319	0.098475	0.930536
X variable 4	6.88196	6.399813	1.075338	0.394726
X variable 5	-0.43857	0.925351	-0.47395	0.68224
X variable 6	-0.08477	66.84155	-0.00127	0.999103

Average prediction difference

$$
\begin{aligned}
D_B &= \frac{(D_{2019} + D_{2020})}{2} \\
&= \frac{\left| \left(Y_{2019\,actual} - Y_{2019\,predicted} \right) + \left(Y_{2020\,actual} - Y_{2020\,predicted} \right) \right|}{2} \\
&= \frac{|(7247 - 8646.91) + (6597 - 9635.64)|}{2} = 2219.28.
\end{aligned}
\tag{21}
$$

(2) Select |Pearson Correlation Coefficient | Variable > 0.78. In the data from 2010 to 2018, the variables selected with | Pearson correlation coefficient | >0.78 are as Table 8.

Multiple regression analysis is carried out between the selected seven variables and house price in Zhaoqing as Table 9.

TABLE 8: |Pearson correlation coefficient | variable > 0.78 between house prices and influencing factors from 2010 to 2018.

Pearson correlation coefficient	House price in Zhaoqing
Proportion of tertiary industry	0.819877
Income of urban residents	0.891052
Employed persons of residents	0.866204
Investment in real estate development	0.956127
Population density	0.824473
Resident population	0.798563
Proportion of urban population	0.787166

TABLE 9: Multiple regression analysis of |Pearson correlation coefficient | variable > 0.78 between house prices and influencing factors from 2010 to 2018.

(a)

Regression statistics	
Multiple R	0.998004
R Square	0.996011
Adjusted R Square	0.96809
Standard error	183.891
Observed Value	9

(b)

Variance analysis	df	SS	MS	Significance F
Regression analysis	7	8443942	1206277	0.128235
Residual error	1	33815.91	33815.91	
Total	8	8477758		

(c)

Regression parameter	Coefficients	Standard error	t Stat	p value
Intercept	-28987	66090.7	-0.43859	0.736867
X variable 1	123.493	112.7199	1.095574	0.470985
X variable 2	0.21588	0.262167	0.823447	0.561449
X variable 3	335.0975	722.1909	0.464001	0.723429
X variable 4	0.114641	16.06871	0.007134	0.995458
X variable 5	-1.87334	3.164806	-0.59193	0.65975
X variable 6	-156.464	331.3167	-0.47225	0.719123
X variable 7	363.5802	744.5641	0.488313	0.710813

Obtain $R^2{}_C = 0.996011$.

$$
\begin{aligned}
Y = {} & (-28987) + 123.49 * X_1 + 0.22 * X_2 + 335.10 * X_3 \\
& + 0.11 * X_4 + (-1.87) * X_5 \\
& + (-156.46) * X_6 + 363.58 * X_7.
\end{aligned}
\tag{22}
$$

TABLE 10: Relationship between house prices and influencing factors from 2010 to 2019.

Pearson correlation coefficient	House price in Zhaoqing
GDP	0.729192
Proportion of secondary industry	-0.73205
Proportion of tertiary industry	0.779115
Income of urban residents	0.909981
Employed persons of residents	0.845897
Fiscal revenue	0.056047
Land price	0.811318
Investment in real estate development	0.883584
Resident population	0.845911
Population density	0.865063
Net migration of population within the province	-0.68724
Net migration of population outside the province	0.555271
Proportion of urban population	0.837435

Substituting the data for 2019 to test the model,

$$
\begin{aligned}
Y_{2019} = {} & (-28987) + 123.49 * 41.68 + 0.22 * 33259.8 \\
& + 335.10 * 231.64 + 0.11 * 456.79 \\
& + (-1.87) * 2045.84 + (-156.46) * 418.71 \\
& + 363.58 * 48.63 = 9493.84.
\end{aligned}
\tag{23}
$$

Substituting the data for 2020 to test the model,

$$
\begin{aligned}
Y_{2020} = {} & (-28987) + 123.49 * 42.7 + 0.22 * 35470.16 \\
& + 335.10 * 235.08 + 0.11 * 586.66 \\
& + (-1.87) * 2343.67 + (-156.46) * 421.71 \\
& + 363.58 * 49.4 = 10526.74.
\end{aligned}
\tag{24}
$$

Average prediction difference

$$
\begin{aligned}
D_C & = \frac{(D_{2019} + D_{2020})}{2} \\
& = \frac{\left|\left(Y_{2019\,actual} - Y_{2019\,predicted}\right) + \left(Y_{2020\,actual} - Y_{2020\,predicted}\right)\right|}{2} \\
& = \frac{\left|(7247 - 9493.84) + (6597 - 10526.74)\right|}{2} = 3088.29.
\end{aligned}
\tag{25}
$$

5.3.2. Adjust the Model with Data of 2010-2019. Because the house prices in Zhaoqing are in an overall increasing trend from 2010-2018, but in a decreasing trend from 2019 to 2020, building the model with the data from 2010 to 2018 may have the problem of large deviation between the predicted and actual values. Therefore, in the process of debugging the model, the data of 2019 was also selected to be included in the prediction model, and the data of 2020 was used as a test.

TABLE 11: |Pearson correlation coefficient | variable > 0.8 between house prices and influencing factors from 2010 to 2019.

Pearson correlation coefficient	House price in Zhaoqing
Income of urban residents	0.909981
Employed persons of residents	0.845897
Land price	0.811318
Investment in real estate development	0.883584
Resident population	0.845911
Population density	0.865063
Proportion of urban population	0.837435

TABLE 12: Multiple regression analysis of |Pearson correlation coefficient | variable > 0.8 between house prices and influencing factors from 2010 to 2019.

(a)

Regression statistics	
Multiple R	0.998274
R Square	0.996551
Adjusted R Square	0.98448
Standard error	136.4757
Observed Value	10

(b)

Variance analysis	df	SS	MS	Significance F
Regression analysis	7	10763747	1537678	0.012019
Residual error	2	37251.25	18625.63	
Total	9	10800998		

(c)

Regression parameter	Coefficients	Standard error	t Stat	p value
Intercept	40965.94	14990.85	2.73273	0.11188
X variable 1	-0.06003	0.048248	-1.24426	0.339447
X variable 2	-442.964	58.9225	-7.51775	0.017238
X variable 3	0.168055	0.117236	1.433472	0.288128
X variable 4	17.10636	1.903533	8.986634	0.012157
X variable 5	173.8367	60.72239	2.86281	0.10343
X variable 6	1.350608	0.396272	3.408283	0.076356
X variable 7	-263.218	260.0871	-1.01204	0.418045

The relationship between house prices (Y variable) and influencing factors (X variable) in Zhaoqing from 2010 to 2019 was calculated using Pearson correlation coefficient as Table 10.

(1) Select |Pearson Correlation Coefficient | Variable > 0.8. In the data from 2010 to 2019, the variables selected with | Pearson correlation coefficient | >0.8 are as Table 11.

TABLE 13: |Pearson correlation coefficient | variable > 0.77 between house prices and influencing factors from 2010 to 2019.

Pearson correlation coefficient	House price in Zhaoqing
Income of urban residents	0.909981
Employed persons of residents	0.845897
Land price	0.811318
Investment in real estate development	0.883584
Resident population	0.845911
Population density	0.865063
Proportion of urban population	0.837435
Proportion of tertiary industry	0.779115

Multiple regression analysis is carried out between the selected seven variables and house price in Zhaoqing as Table 12.

Obtain $R^2{}_D = 0.996551$.

$$
\begin{aligned}
Y = {}& 40965.94 + (-0.06) * X_1 + (-442.96) * X_2 + 0.17 * X_3 \\
& + 17.11 * X_4 + 173.84 * X_5 + 1.35 * X_6 \\
& + (-263.22) * X_7.
\end{aligned}
\tag{26}
$$

Substituting the data for 2020 to test the model,

$$
\begin{aligned}
Y_{2020} = {}& 40965.94 + (-0.06) * 35470.16 + (-442.96) * 235.08 \\
& + 0.17 * 2081 + 17.11 * 586.66 + 173.84 * 421.71 \\
& + 1.35 * 2343.67 + (-263.22) * 49.4 = 8569.17.
\end{aligned}
\tag{27}
$$

Prediction difference

$$
D_D = \frac{\left| Y_{2020\,\text{actual}} - Y_{2020\,\text{predicted}} \right|}{2} = |6597 - 8569.17| = 1972.17.
\tag{28}
$$

(2) Select |Pearson Correlation Coefficient | Variable > 0.77. In the data from 2010 to 2019, the variables selected with | Pearson orcrelation coefficient | >0.77 are as Table 13.

Multiple regression analysis is carried out between the selected eight variables and house price in Zhaoqing as Table 14.

Obtain $R^2{}_E = 0.999005$.

$$
\begin{aligned}
Y = {}& 79871.91 + .(-0.14) * X_1 + (-678.10) * X_2 + 0.36 * X_3 \\
& + 24.45 * X_4 + 213.62 * X_5 + 2.42 * X_6 \\
& + (-284.37) * X_7 + (-80.48) * X_8.
\end{aligned}
\tag{29}
$$

TABLE 14: Multiple regression analysis of |Pearson correlation coefficient| variable > 0.8 between house prices and influencing factors from 2010-to 2019.

(a)

Regression statistics	
Multiple R	0.999502
R Square	0.999005
Adjusted R Square	0.991044
Standard error	103.676
Observed Value	10

(b)

Variance analysis	df	SS	MS	Significance F
Regression analysis	8	10790249	1348781	0.068939
Residual error	1	10748.71	10748.71	
Total	9	10800998		

(c)

Regression parameter	Coefficients	Standard error	t Stat	p value
Intercept	79871.91	27268.89	2.929049	0.209448
X variable 1	-0.14071	0.063111	-2.22952	0.268417
X variable 2	-678.097	156.2903	-4.3387	0.144212
X variable 3	0.364176	0.1534	2.374033	0.2538
X variable 4	24.45209	4.896498	4.993792	0.125818
X variable 5	213.6202	52.6286	4.059013	0.153779
X variable 6	2.41665	0.742653	3.254079	0.189806
X variable 7	-284.368	198.0379	-1.43593	0.387266
X variable 8	-80.4793	51.25292	-1.57024	0.361009

Substituting the data for 2020 to test the model,

$$
\begin{aligned}
Y ={}& 79871.91 + (-0.14) * 35470.16 + (-678.10) * 235.08 \\
&+ 0.36 * 2081 + 24.45 * 586.66 + 213.62 * 421.71 \\
&+ 2.42 * 2343.67 + (-284.37) * 49.4 \\
&+ (-80.48) * 42.7 = 8864.33.
\end{aligned}
\tag{30}
$$

Prediction difference

$$
D_E = \frac{\left| Y_{2020\,actual} - Y_{2020\,predicted} \right|}{2}
\tag{31}
$$
$$
= |6597 - 8864.33| = 2267.33.
$$

5.4. Select a Better Prediction Model. From the above data analysis, the regression coefficients of the five models A, B, C, D, and E were obtained as: $R^2_A = 0.99506$, $R^2_B = 0.99506$, $R^2_C = 0.996011$, $R^2_D = 0.996551$, and $R^2_E = 0.999005$; the

TABLE 15: Multiple regression analysis of |Pearson correlation coefficient| variable > 0.8 between house prices and influencing factors from 2010 to 2020.

(a)

Regression statistics	
Multiple R	0.938207
R Square	0.880232
Adjusted R Square	0.600773
Standard error	673.9447
Observed Value	11

(b)

Variance analysis	df	SS	MS	Significance F
Regression analysis	7	10014408	1430630	0.187273
Residual error	3	1362604	454201.4	
Total	10	11377013		

(c)

Regression parameter	Coefficients	Standard error	t Stat	p value
Intercept	34843.03	73941	0.471227	0.669628
X variable 1	-0.01495	0.236792	-0.06314	0.953627
X variable 2	11.72406	8.856228	1.323821	0.277397
X variable 3	-427.751	290.8349	-1.47077	0.237718
X variable 4	0.99716	1.945903	0.512441	0.643685
X variable 5	166.7663	299.8308	0.556201	0.616866
X variable 6	-135.989	1282.201	-0.10606	0.92223
X variable 7	0.331915	0.570935	0.581354	0.601801

absolute values of the differences between actual and predicted house prices used to test the five models are as follows: $D_A = 2303.46$, $D_B = 2219.28$, $D_C = 3088.29$, $D_D = 1972.17$, and $D_E = 2267.33$.

Compare the regression coefficients of the five models: $R^2_E = 0.999005 > R^2_D = 0.996551 > R^2_C = 0.996011 > R^2_A = R^2_B = 0.99506$.

Compare the absolute value of the difference between actual and predicted house prices for the five models: $D_D = 1972.17 < D_A = 2303.46 < D_B = 2267.33 < D_E = 2245.13 < D_C = 3088.29$.

Since the larger the regression coefficient R^2, the better the model fit; the smaller the absolute value of the difference between the actual and predicted house prices, the better the prediction. Combining the two, we can choose the relatively better one model to predict house prices. Because the regression coefficients of the five models are all above 0.995, the regression coefficients are all at a high level, and the fitting effect is better, and the absolute value of the difference between the actual house price and the predicted house price of model D is the smallest, so model D is chosen to predict

the house price, that is,

$$Y = 40965.94 + (-0.06) * X_1 + (-442.96) * X_2 + 0.17 * X_3 \\ + 17.11 * X_4 + 173.84 * X_5 + 1.35 * X_6 \\ + (-263.22) * X_7,$$

(32)

where X_1 is income of urban residents, X_2 is employed persons of residents, X_3 is land price, X_4 is investment in real estate development, X_5 is resident population, X_6 is population density, and X_7 is proportion of urban population.

In summary, model D was screened as the final model to predict the house prices in Zhaoqing by comparing the absolute values of the regression coefficients R^2 and the differences between the actual and predicted house prices. The final multiple regression model is derived based on the full data from 2010 to 2020 to predict the house prices in Zhaoqing in 2021 and beyond as Table 15.

The model for predicting house prices in Zhaoqing was obtained as follows:

$$Y_{2021} = 34843.03 + (-0.015) * X_1 + 11.724 * X_2 \\ + (-427.751) * X_3 + 0.997 * X_4 \\ + 166.766 * X_5 + (-135.990) * X_6 + 0.332 * X_7,$$

(33)

where X_1 is income of urban residents, X_2 is employed persons of residents, X_3 is land price, X_4 is investment in real estate development, X_5 is resident population, X_6 is population density, and X_7 is proportion of urban population.

5.5. Application from Research Conclusions to Practice. This model can be applied to real life. For example, we can collect the data of income of urban residents, employed persons of residents, land price, investment in real estate development, resident population, and population density and proportion of urban population in the first half of the year and calculate the average and their value of the whole year. In this way, we can predict the house price of Zhaoqing at the end of the year through this model.

Governments, industries, and individuals can apply this model to real life. If home values continue to decrease, it is likely to have a detrimental influence on the local economy, and the government may prepare for a drop in real estate tax income. At the same time, the home price prediction model may alert the government to the disparity between house prices and residents' income. It can also encourage economic growth to close the gap and enhance residents' quality of life. Stable property prices can help Zhaoqing's economy. Also, house price prediction models help customers buy or invest in properties. The model can estimate future housing values, their size, and the rate of return on investment in real estate. Based on this model, consumers may decide if it is the proper time to buy a home for an immediate need, and locals can decide whether to buy a local home that year.

6. Conclusions and Future Works

6.1. Conclusions. According to the research methods of machine learning and statistics, the method in this thesis can use the method of correlation analysis with the help of variables such as GDP, the proportion of secondary industry, the proportion of tertiary industry, urban residents' income, residents' employed persons, fiscal revenue, land price, the amount of investment in real estate development, resident population, population density, net migration of population within the province, net migration of population outside the province, and the proportion of urban population. Based on the correlation coefficients, the variables with high correlation with the house price data are selected and multiple linear regression analysis is performed to build a model to realize the prediction of house prices. Based on the available data, we select data with different correlations to continuously adjust the model to make the fitting effect better and the prediction effect more accurate and select the best prediction model.

Through the research process described above, the model derived in this paper for forecasting house prices in 2021 and beyond is as follows:

$$Y_{2021} = 34843.03 + (-0.015) * X_1 + 11.724 * X_2 \\ + (-427.751) * X_3 + 0.997 * X_4 + 166.766 * X_5 \\ + (-135.990) * X_6 + 0.332 * X_7.$$

(34)

In the equation, X_1 expresses income of urban residents, X_2 expresses employed persons of residents, X_3 expresses the land price, X_4 expresses the investment in real estate development, X_5 expresses resident population, X_6 expresses population density, and X_7 expresses the proportion of urban population.

However, from a practical forecasting perspective, the unquantifiability of many influencing variables, such as policies, makes the accurate prediction of house prices still difficult, but it is undeniable that our model can give good guidance suggestions.

In the prediction model, there is an incremental increase of various indicators in 2019 and 2020, but the decreasing trend of house prices, making the forecast value and the actual value of the large deviation, which is more related to China's policy of restricting house prices.

The Chinese government has introduced a series of policies to regulate and limit the growth of house prices, such as preventing the expansion of real estate investment and land use, by reforming the land grant, financial management, and investment system. In order to prevent government land from being bought at low prices and sold at high prices or borrowing heavily from banks, starting from the bank credit system, commercial banks raise the requirements for free capital and qualifications of real estate developers, should strengthen the examination of real estate developers when lending to open houses, and should restrict loans to construction companies with low own capital and high accounts receivable. Strictly control the financing supervision of real

estate developers and increase the capital ratio of real estate developers.

In that research, in addition to giving models with implications for guiding house price forecasting, we also gained the following insights. When predicting house prices in a region, researchers should focus not only on quantifiable model data, but also on unquantifiable influencing factors, such as policy and social factors. Especially in countries with more government intervention in the market economy, policy factors have a greater impact on house prices and may exceed the extent to which economic factors affect house prices.

6.2. Future Works. From the modeling perspective, the unquantifiable factors in our research failed to be put into the house price prediction model, and these unquantifiable factors, such as policies, happen to be the factors that have a great influence on house prices, and adding unquantifiable factors can improve the prediction accuracy of the model to some extent. Due to the influence of unquantifiable factors, there may be some deviation between the final prediction results and the actual results. From the general trend of house prices in Zhaoqing given in this thesis, it is clear that the house prices in Zhaoqing have reached an inflection point in 2018, from an upward trend to a downward trend, which is strongly related to the Chinese government's policy of "houses are used for living and not for speculation" and the policy of limiting the loan ratio for the second suite, that is, the down payment ratio should not be less than 40%. These policies restrict with a part of people to buy homes in Zhaoqing, which makes the demand and price of Zhaoqing homes become lower. In future research, the problem of quantifying nonquantifiable factors and how to include nonquantifiable factors in the model to make the forecast results more accurate and good research directions.

The time lag of the policy will also be a good direction for future research. The Chinese government implemented the purchase restriction policy in Beijing in 2010, put forward the concept of "houses are for living and not for speculation" in 2016, and put forward the second suite purchase restriction policy in 2017, while the house price in Zhaoqing only started to decline in 2019, before that it was in an upward trend, which shows that the policy has a certain lag on real estate prices. This will be a good direction for future research.

In addition to the multivariable linear regression model in machine learning used in this research, the logistics model and nonlinear regression model in machine learning are also good methods for predicting housing prices.

The essence of logistic regression is to divide the probability of occurrence by the probability of no occurrence and then take the logarithm. That is, this less cumbersome transformation changes the contradiction of the taking value interval and the curvilinear relationship between dependent variables and independent variables. The reason is that the probability of occurrence and nonoccurrence becomes a ratio which is a buffer, expanding the value range, and then performing logarithmic transformation, the entire dependent variable changes. More than that, this transformation often leads to a linear relationship between dependent variables and independent variables, which is summarized based on a lot of practice. Therefore, logistic regression basically solves the problem that dependent variables are not continuous variables. Also, logistic is widely used because many real-world problems fit its model. For example, whether an event occurs is related to other numerical independent variables. Therefore, logistic regression can also be used to predict housing prices based on various variables in reality [26].

The nonlinear model is a mathematical expression that reflects the nonlinear relationship between the independent variable and the dependent variable. Compared with the linear model, the dependent variable and the independent variable cannot be expressed as a linear correspondence in the coordinate space. Distinguishing whether it is a linear model is mainly based on the coefficient w before the independent variable x in a multiplication formula. If w affects only one x, then the model is a linear model. For example, in formula (35), y and x are curvilinear relationship, but it is a linear model, because it can be observed that x_1 is only affected by one w_1 in $x_1 * w_1$.

$$y = \frac{1}{1 + e^{w_0 + w_1 * x_1 + w_2 * x_2}}, \tag{35}$$

while formula (36) is a nonlinear model. Because it is observed that x_1 is not only affected by the w_1 parameter, but also by w_5. If the independent variable x is affected by more than two parameters, then this is nonlinear model.

$$y = \frac{1}{1 + w_5 * e^{w_0 + w_1 * x_1 + w_2 * x_2}}. \tag{36}$$

In many cases, it can be found that the linear model cannot fit well, resulting in distorted data and even forced linear regression. As a result, it not only changed the normality of the original data but even changed the equal variance and independence of the data [27]. At this time, the introduction of nonlinear regression models should be taken into account, such as neural networks in machine learning, SVR, decision tree regression, and KNN regression. The use of these models to predict housing prices is also a good research direction in the future.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] W. S. Petty, *A Treatise of Taxes and Contributions*, Good Press, 2021.

[2] A. Marshall, *Principles of Economics: Unabridged*, Cosimo, Inc., 8th edition, 2009.

[3] T. R. Malthus, D. Winch, and P. James, *An Essay on the Principle of Population*, Cambridge University Press, 1992.

[4] N. Ghosalkar and S. N. Dhage, "Real estate value prediction using linear regression," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–5, Pune, India, 2018.

[5] A. Kaushal and A. Shankar, "House prrice prediction using multiple linear egression (April 25, 2021)," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2021.

[6] S. Amri and G. A. Tularam, "Performance of multiple linear regression and nonlinear neural networks and fuzzy logic techniques in modelling house prices," *Journal of Mathematics and Statistics*, vol. 8, no. 4, pp. 419–434, 2012.

[7] R. B. Abidoye, A. P. C. Chan, F. A. Abidoye, and O. S. Oshodi, "Predicting property price index using artificial intelligence techniques: evidence from Hong Kong," *International Journal of Housing Markets and Analysis*, vol. 12, no. 6, pp. 1072–1092, 2019.

[8] S. Fong and Y. B. Wah, "A prediction model for forecasting the trend of Macau property price movements and understanding the influential factors," *Journal of Emerging Technologies in Web Intelligence*, vol. 5, no. 2, pp. 122–131, 2013.

[9] G. Hongbin, *The Analysis and Forecast for House Prices in Shenzhen, [M.S. thesis]*, Dalian University of Technology, 2021.

[10] Y. Bilei, *The Effects of Com Mercial Residential Honors on the Housing Price in Guangzhou, [M.S. thesis]*, South China University of Technology, 2013.

[11] N. G. Mankiw, *Principles of Economics*, Cengage Learning, 2020.

[12] A. C. Pigou, *A Study in Public Finance*, Read Books Ltd., 2013.

[13] R. Inglehart, *Culture Shift in Advanced Industrial Society*, Princeton University Press, 2018.

[14] G. J. Borjas and J. C. Van Ours, *Labor Economics*, McGraw-Hill/Irwin, Boston, 2010.

[15] E. P. Lazear and M. Gibbs, *Personnel Economics in Practice*, John Wiley & Sons, 2014.

[16] J. Harvey, *Urban Land Economics*, Macmillan International Higher Education, 2016.

[17] W. B. Brueggeman and J. D. Fisher, *Real Estate Finance and Investments*, McGraw-Hill Irwin, New York, 2011.

[18] J. L. Simon, *The Economics of Population Growth*, Princeton University Press, 2019.

[19] H. Tisdale, "The Process of Urbanization," *SOC-F*, vol. 20, p. 311, 1941.

[20] Information Center of The Ministry of Culture and Tourism of China, *Analysis Report on Online Evaluation Data of 5A Scenic Spots in 2019*, China Tourism Press, 2019.

[21] T. Shi, *Research on the Optimization of Bus Service in Zhaoqing City in the New Era*, South China University of Technology, 2020.

[22] Meteorological Administration of Zhaoqing and Zhaoqing Meteorological Bulletin, *Zhaoqing Development and Reform Bureau Press*, 2019.

[23] P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Psychology press, 2014.

[24] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, springer, New York, 2013.

[25] C. Dougherty, *Introduction to econometrics*, Oxford university press, 2011.

[26] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398, John Wiley & Sons, 2013.

[27] L. Grüne, J. Pannek, L. Grüne, and J. Pannek, *Nonlinear Model Predictive Control. In Nonlinear Model Predictive Control*, Springer, Cham, 2017.