Research on the Influencing Factors of House Prices based on Multiple Linear Regression Model

Ruibin Li

School of Life Sciences, Shandong University, Qingdao, Shandong, 266200, China

202300141155@mail.sdu.edu.cn

Abstract. House prices have risen rapidly, and finding the factors affecting house prices is significant for regulating house prices. Existing literature discusses the factors affecting house prices, but does not analyze the key factors affecting house prices and get a good fit with a multiple linear regression model. Therefore, this paper is of practical significance for the selection and fitting of key variables that explain the rise in house prices. In Spearman's correlation analysis, strong correlation and covariance diagnostics were found for gross product. After linear fitting, it is found that gross product, commercial real estate sales area will have an impact on house prices. The results of this paper on the impact of GDP and commercial real estate sales area on housing prices are conducive to the relevant agencies to take measures to adjust housing prices. At the same time, for people who have the need to buy a house, understanding the main influencing factors is conducive to consumers' initial judgment on the trend of house prices, and then make a decision.

Keywords: Multiple linear regression model, covariance, GDP, sale areas.

1. Introduction

Clothing, food, housing and transportation, are necessities that need to be considered at any time. They are issues of people's livelihood, which are closely related to the stability of the society and the prosperity of the country. After the basic realization of comprehensive well-off, the problem of food and clothing has been well solved, and travel has become more and more convenient, however, buying a house, housing is still a problem that plagues many Chinese people. Between 2014 and 2022, the average sales price of commercial housing has risen from just over six thousand dollars to an average price of over 10,000 dollars. For similar Beijing, Shanghai, Shenzhen, such as the average price of first-tier cities, can reach more than 40,000/flat, part of the region can reach more than 100,000 home prices, while the national urbanization unit 22 years of the average income of 110,000 at the beginning. That is to say, in China, most people are not able to purchase a set of houses in their long-term living place to meet the needs of normal life. Such higher and higher housing prices have received the attention of the state. In response, the state has also implemented a series of policies to control housing prices. Since 2015 to the present, the state has mainly promoted the reform of residential leasing and achieved good results, and for the acquisition of housing, it mainly implements the policy of restricting the purchase [1].

In the current status of foreign research, Capozza et al. explored the dynamics of house prices by studying the mean regression coefficients of panel data for 62 regions from 1979 to 1995, and found that population growth as well as income growth would bring about the growth of house prices [2].

Some scholars have studied the factors affecting house prices based on the supply and demand framework, and based on the supply and demand theory, it is concluded that the factors affecting house prices are income and loan profits [3]. Balazs and Dubravko conducted a study on the determinants of house prices in eight economies in transition and 19 Organization for Economic Cooperation and Development (OECD) countries in the Central Eastern Europe (CEE) region, using traditional fundamentals such as GDP per capita, real interest rates, housing credit, and population, as well as some specific transitory factors, such as institutional development of the housing market, housing finance, and quality effects, as explanatory variables, and empirically analyze quarterly data from 27 countries using panel model [4].

While in the current status of domestic research, Li has found that the level of per capita disposable income and monetary policy are the main factors affecting real estate prices, and the direction of influence is in the same direction and significant [5]. Zhang considered the factors affecting house prices from both supply and demand, and then uses the traditional statistical method of multiple linear regression to model house prices in Suzhou, concluding that the predicted value in 2013 is very close to the true value [6]. Wang et al. took most of the provincial capital cities in China as an example, and analyzes the correlation between urban housing prices and population, GDP and other influencing factors between 2000 and 2014, and comes up with the highest influencing factor affecting housing prices. Then the cities are categorized based on the highest influence factor to provide a basis for formulating relevant policies [7]. Li used the gray-Markov prediction model and the polynomial model to model and analyze the national average house price data over the years, and the results show that the polynomial model has a good accuracy in predicting house prices [8]. Li took China's east, middle and west as the sample region of house price, and conducts research in urban population, residents' income, per capita social output, etc., and argues that the growth value of urban population size has an obvious enhancement effect on house price in the east and middle regions, the increase of urban residents' income is positively correlated with the house price, the per capita social output of urban residents has an obvious influence only on the middle and west regions, and the influence of real estate development investment in the local regions forms a difference [9]. Xu and Huang took the residential sales price index data from the official website of the National Bureau of Statistics as the explanatory variables, and select the Internet search index of 8 keywords related to new housing prices as the explanatory variables, establish stepwise regression, fully quadratic polynomial and multiple linear regression models of both, and finally get the stepwise regression model has better effect [10]. Lian built the index system of house price from macro and micro aspects respectively, and analyzes the influencing factors of house price in Wuhan by using various regression methods such as vector autoregressive model and geographically weighted regression [11].

The research on the influencing factors of house prices helps us to determine the directions that need to be considered to study the trend of house price changes, but how to utilize these directions, as well as the use of several directions to study the trend of house price changes in the multiple linear regression model to get a good fitting effect, has certain research significance.

In this study, based on the average sales price of commercial housing in Shanghai from 2005 to 2022, and nine independent variables, the article tries to predict the house price in 2023, to reveal the degree of contribution of the selected influencing factors to the house price, and the fit of the multiple linear regression model. In this study, the influencing factors are refined, and the average sales price of commercial houses is taken as the dependent variable, and the GDP, development investment, consumer price index, year-end resident population, annual per capita income, sales area of commercial houses, area of residential houses completed in the whole society, residential investment in real estate development, and local financial tax revenue are taken as the control variables, and the data of the nine-year period of 2014-2022 are selected for fitting of the multiple linear regression model.

2. Methods

2.1. Data Source and Description

Since data on house prices are scarce and hard to find, to ensure the authenticity of the data, the data used in this study come from the National Bureau of Statistics, National Data and other websites. The dataset contains data on the average sales price of commercial housing in Shanghai, tax revenue (billion yuan), resident population at the end of the year (10,000), annual per capita income (yuan), investment in real estate development (billion yuan), consumer price index, area of sales of commercial housing (10,000 square meters), area of residential housing completed by the whole society (10,000 square meters), gross domestic product (billion yuan), and investment in residential housing by the whole society (billion yuan), etc. The data are used for estimating the price of commercial housing. The data are used to estimate the degree of influence of control variables such as GDP and per capita income, and then the model is used to predict and analyze the house price. Based on the information available from the National Bureau of Statistics (NBS), the data for Shanghai was selected for the nine-year period from 2005 to 2022.

2.2. Indicator Selection and Explanation

Firstly, the factors affecting house prices are categorized into administrative factors, economic factors, demographic factors and social factors. Administrative and house prices are directly related to the local financial tax revenue, economic aspects of the selection of direct indicators related to real estate, such as gross regional product, real estate development investment, per capita income, consumer price index, demographic factors for the end of the year resident population, social indicators for the sales area of commercial properties, the social housing area, social housing investment, and so on. The control variables will be selected from the above nine indicators, and the number of selected indicators will not exceed nine. The following table 1 shows the basic description about some independent and dependent variables.

Table 1. Description of quantitative variables.

Items	Min	Max	Mean	SD
y (Average price)	8195.00	40302.00	21437.54	9907.61
x1(Development investment)	1307.53	5035.18	3178.74	1270.31
x2 (Resident population)	2064.00	2489.00	2386.69	135.91
x3 (Per capita annual income)	22459.00	79610.00	49615.94	19196.91
x4 (GDP)	12878.70	44809.10	27624.16	10454.56
x5(CPI)	99.60	105.80	102.59	1.46
x6 (Local tax revenues)	1975.48	6606.74	4471.22	1670.04
x7 (Completed residential)	934.69	2752.45	1613.64	369.03
x8 (Sales area)	1691.60	3694.96	2214.90	597.21
x9 (Residential investment)	837.53	2771.80	1778.68	624.02

With a basic description of the independent variable, next are some in-depth descriptions of the independent variable to better characterize the data.

3. Results and Discussion

3.1. Descriptive Analysis

First, this paper provides a rough description of the dependent variables explored in the paper, as shown in Figure.
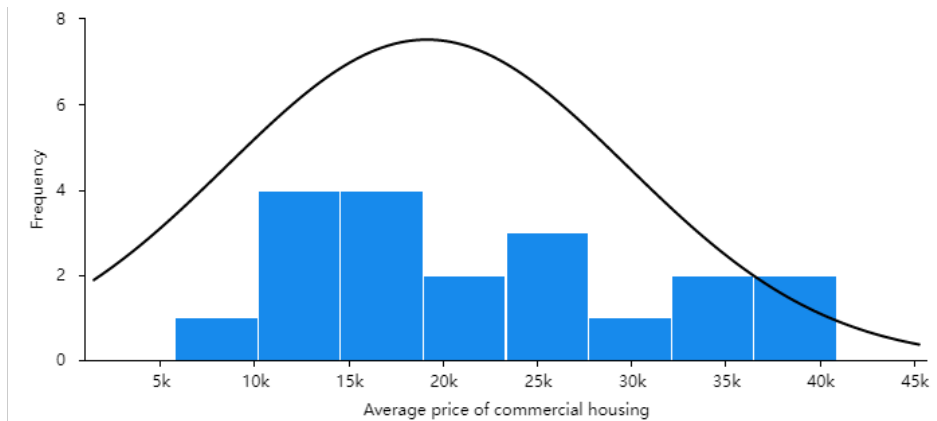
Figure 1. Distribution of house price.

As can be seen in Figure 1, the distribution of house prices in the nine years between 2014-2022 is relatively balanced and can be approximated as obeying a normal distribution. In addition, in order to better characterize the data distribution of the average price of commercial housing, a box plot is taken to describe it, as shown in Figure 2.
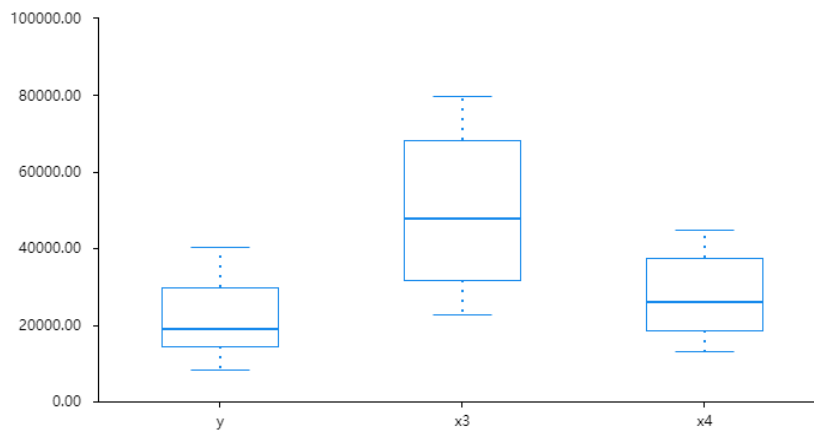

Figure 2. box-plot of main variables.

In Figure 2, the minimum of the house price in Shanghai between 2014-2022 is 16,787.00 yuan, the maximum is 40,302.00 yuan, and the median is about 26,890.08 yuan, and the overall distribution presents a balanced character.

3.2. Correlation Analysis

According to the given 9 indicators need to analyze the dependent and independent variables, and the correlation between the independent variables, for the strong correlation between the independent variables, may have multiple contributions to the dependent variable, not applicable to multiple linear regression model. Figures 3 and 4 show the correlation analysis between the dependent and independent variables and the partial correlation analysis between the independent variables and the independent variables, respectively.
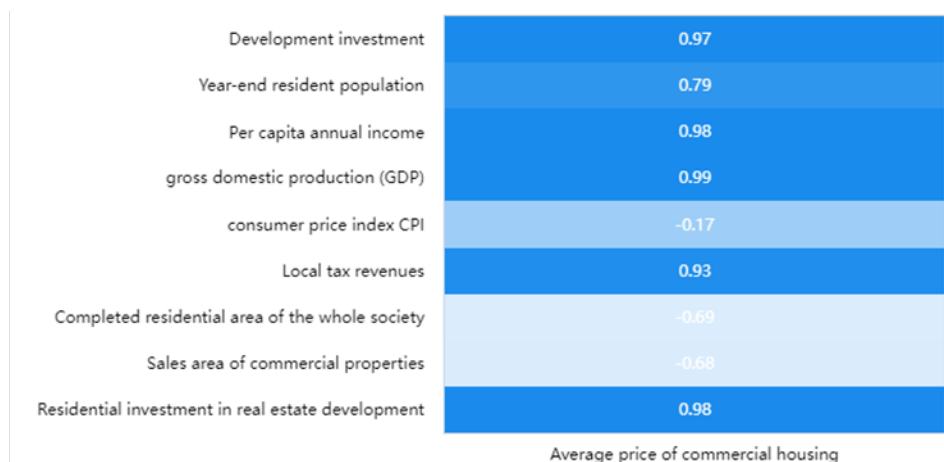
Figure 3. Correlation analysis.

Figure 3 illustrates that there is a positive correlation between house prices and GDP and the amount of development investment, annual per capita income, resident population at the end of the year, the amount of investment in residential development, and local tax revenues.
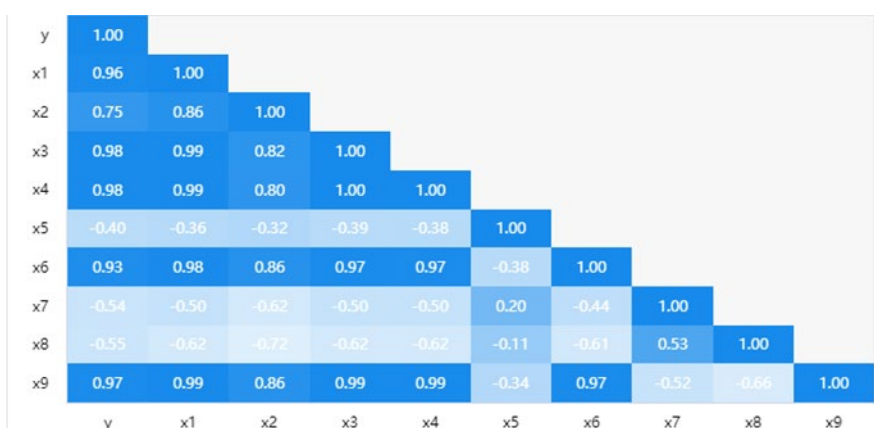

Figure 4. Pearson Correlation-delta format.

As can be seen from Figure 4, GDP and annual per capita income have strong correlations with the amount of residential investment and development investment. Therefore, when fitting, it is necessary to select the main components in the amount of development investment, residential investment, per capita annual income, GDP, and then, based on this, select the combination of independent variables that have a clear relationship with the dependent variable for optimal fitting. The model needs to select the independent variables that have strong correlation with the dependent variable and are relatively independent to be fitted, therefore a covariance diagnostic is needed to screen the independent variables for model fitting. The covariance diagnostics for the nine indicators are shown in the table 2 below.

Table 2. Diagnosis of covariance.

Items	VIF value	Tolerance
consumer price index CPI	4.831	0.207
gross domestic production (GDP)	3739.764	0.000
Sales area of commercial properties	7.496	0.133
Completed residential area of the whole society	7.780	0.129
Development investment	224.548	0.004

Table 2. (continued).

Local tax revenues	76.683	0.013
Year-end resident population	102.370	0.010
Per capita annual income	1757.533	0.001
Residential investment in real estate development	748.332	0.001

Note: Red underlining indicates VIF >10 or null.

As can be seen from Table 2, the VIFs of CPI, residential completed area, and sales area are less than 5 and relatively independent. These three variables can be selected when performing the fitting. Meanwhile, among the six variables with multicollinearity, the independent variable with the most significant correlation with the dependent variable is selected for the final fitting.

3.3. Model Results

This paper chose a multiple linear regression model for data analysis, and in order to ensure the accuracy of the results, the paper chose to fit different combinations based on GDP and the remaining three relatively independent variables. Finally, this paper chose GDP and sales area for fitting, and the results are shown in Table 3 and 4.

Table 3. Parameter Estimates

	Unstandardized Coefficients		Standardized Coefficients	t	p	Covariance diagnosis	
	B	S.E.	Beta			VIF	Tol
Constant	140447.862	124713.138	-	1.126	0.303	-	-
x1	-0.506	4.618	-0.065	-0.110	0.916	224.548	0.004
x2	-41.149	29.143	-0.564	-1.412	0.208	102.370	0.010
x3	0.906	0.855	1.755	1.059	0.330	1757.533	0.001
x4	-1.793	2.290	-1.892	-0.783	0.464	3739.764	0.000
x5	-485.139	591.418	-0.071	-0.820	0.443	4.831	0.207
x6	0.669	2.053	0.113	0.326	0.755	76.683	0.013
x7	-5.904	2.959	-0.220	-1.996	0.093	7.780	0.129
x8	0.593	1.795	0.036	0.330	0.752	7.496	0.133
x9	22.706	17.161	1.430	1.323	0.234	748.332	0.001
R^2	0.991						
Adj R^2	0.977						
F	F (9,6) =70.504, p=0.000						
D-W	1.842						

From the table, it can be seen that the analysis of the full variables produces multicollinearity, which in turn causes none of the variables to be significant, while the combination chosen in Table 4 has a good fit. The model equation is: Average price of commercial housing = -9510.734 + 0.970*gross domestic production (GDP) + 1.954*Sales area of commercial properties, and the model R-squared value is 0. The R-squared value of the model is 0. 979, which means that gross domestic production (GDP), sales area of commercial properties can explain 97. 9% of the variation in Average price of commercial housing. The regression coefficient value of gross domestic production (GDP) is 0.970 (t=19.925, p=0.000<0.01), which implies that gross domestic production (GDP) has a significant positive effect on Average price of commercial housing.

Table 4. Parameter Estimates after variable selection

	Unstandardized Coefficients		Standardized Coefficients	t	p	covariance diagnosis	
	B	Std. Error	Beta			VIF	Tolerance
Constant	-9510.734	3100.659	-	-3.067	0.007	-	-
GDP	0.970	0.049	1.075	19.925	0.000	2.226	0.449
Sales area	1.954	0.878	0.120	2.225	0.041	2.226	0.449
R^2			0.979				
Adj R^2			0.976				
F			F (2,16) =374.065, p=0.000				
D-W value			1.217				

The value of regression coefficient of Sales area of commercial properties is 1.954 (t=2.225, p=0.041<0. 05) implies that Sales area of commercial properties will have a significant positive impact on Average price of commercial housing. gross domestic production (GDP), Sales area of commercial properties will have a significant positive impact on Average price of commercial housing. Sales area of commercial properties all have significant positive effect on Average price of commercial housing (Table 4).

3.4. Discussion
In short, Gross Regional Product (GRP) and sales area of commercial real estate will have an impact on the price of commercial real estate. From the analysis, it can be seen that GDP and real estate development investment, annual per capita annual income, real estate development residential investment, local fiscal tax revenue and other variables have a correlation of varying sizes. On the basis of previous studies that have analyzed the influencing factors of house prices, this paper more specifically reflects the linear relationship between house prices and GDP and the sales area of commercial properties under the multiple linear regression model, as well as the covariance relationship within some of the commonly thought of house price influencing factors, such as GDP and development investment, per capita annual income, resident population, and tax revenues. This paper provides a new way of thinking about the influences on house prices, guiding the focus on house price changes in terms of local GDP and sales area. In the next study, this paper will further explore the factors related to house prices, reveal their more specific relationships, and lay the foundation for better fitting using multiple linear regression models. In this paper, the author will further explore the factors related to house prices to reveal their more specific relationships and lay the foundation for better fitting using multiple linear regression models.

4. Conclusion
This paper explores the correlation between house prices and a series of variables of interest, such as gross domestic product, commercial property sales area, resident population, and development investment, and gives multiple linear regression models for analysis. It is found that gross domestic production (GDP), Sales area of commercial properties can explain 97.9% of the variation in Average price of commercial housing. The house price has a significant effect with Gross domestic production (GDP), Sales area of commercial properties. At the same time, this paper utilizes covariance diagnostics and finds that GDP has covariance with development investment, annual per capita income, residential investment, and even tax revenue and resident population. Based on the findings of this paper, individuals can make judgment and prediction on the trend of house price based on GDP, sales area of commercial properties and other variables that people are concerned about. In addition, the results of this paper are helpful for local real estate policy and tax policy making. Ultimately, the results of this paper are consistent with the theme of the study and have certain reference value and practical significance.

References

[1] Zhou, K. (2023) Research on the house price effect of China's real estate purchase restriction policy. Shanghai University of Finance and Economics.

[2] Capozza, R.D., Hendershott, H.P. and Mack, C. (2024) An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets.Real Estate Economics, 32(1), 1-32.

[3] Holly, S., Pesaran, H.M. and Yamagata, T. (2010) A spatio-temporal model of house prices in the USA. Journal of Econometrics, 158(1), 160-173.

[4] Balazs, E. and Dubravko, M. (2007) Determinants of House Prices in Central andEastern Europe. BIS Working Paper Series, 236.

[5] Li, Y.M. (2012) An empirical study on the characteristics of real estate price changes and its influencing factors in China. Jilin University.

[6] Zhang, J. (2014) Analysis of Influencing Factors of Commodity Housing Prices and Forecast of House Prices in Suzhou. Suzhou: Suzhou University.

[7] Wang, L., Chen, B.L. and Feng, Y.Y. (2016) Correlation analysis of house price and its influencing factors in major cities of China. Modern Economic Information, 31(37), 32-37.

[8] Li, D.Y. (2006) Comparative study of house price prediction models. Industrial Technology Economy, 25(9), 65-67.

[9] Li, Y. (2018) Comparative study on the influencing factors of commercial housing prices. Comparison of economic and social systems, 2, 20-31.

[10] Xu Y. and Huang, S.Z. (2017) Research on real estate price prediction model in the era of big data. China Business Journal, 4, 134-137.

[11] Lian, Z.Y. (2021) Research on spatial and temporal evolution characteristics and influencing factors of house prices in Wuhan. Jiangxi University of Technology.