Analysis of the Factors Influencing Housing Prices in Yunnan and Guizhou Based on Linear Regression

Yuan Fang

The University of Sydney, Sydney 2006, Australia

yfan6884@uni.sydney.edu.au

Abstract. The real estate industry contributes significantly to the national economy. In reality, it has a significant impact on the lives of residents. It is of great significance to study the factors that affect the housing price for the development of the real estate industry. Based on the real estate data of Yunnan Province and Guizhou Province from January 2012 to December 2022, this study will analyze the factors influencing the housing prices of Yunnan Province and Guizhou Province. In addition to SPSS research, multiple regression analysis will be used in this paper. According to the analysis, housing prices in Yunnan and Guizhou provinces will rise with the decline of long-term loan interest rates. Besides, the residential consumer price index negatively affects housing prices in Yunnan and Guizhou provinces. Moreover, an increase in the amount of real estate investment will lead to an increase in housing prices. On this basis, the completed real estate area has a negative relationship with house prices.

Keywords: Housing prices, Linear Regression, Yunnan, Guizhou.

1. Introduction

Real estate is essential to the national economy, especially in China. Because in Chinese traditional culture, it is important for people to own their real estate. The stability of a country's development depends on the policymakers in that nation knowing what elements will have an impact on real estate. For companies and individuals, knowing information about the real estate market is unavoidable. Knowing this information could help them to make reasonable plans to buy or sell their house appropriately [1]. Therefore, researching the influencing factors of housing prices is meaningful.

There are lots of scholars focused on researching this area around the world. Real estate could be classified as a kind of commodity. It is influenced by the supply and demand in the market [2]. Under this circumstance, some factors related to supply and demand will influence the price of real estate. Specifically, there are several kinds of demands, such as self-use demand, investment demand and speculative [3]. Liu found the main factor influencing the price of real estate is the income of residents by using multiple linear regression [4]. It seems sensible that people would spend more on real estate if they were able to increase their income in the future.

Moreover, some researchers make predictions about the trends of property prices in the future, based on statistics from specific cities all around the world. Chen used the multiple linear regression analysis and some data related to Zhaoqing's real estate market to create the model to predict the price of real estate in the future [5]. Certainly, it is significant to find some factors that would influence housing price in different areas. Javad and Meysam created a model to forecast apartments' prices by using the linear regression method [6]. As for independent variables, they focused more on the nature of apartments, such as the multiplication of elevator in floor, parking, and number of floors. By contrast, most researchers find independent variables rely on demand and supply.

In addition, Jian applied general linear regression to analyse some factors that could influence the price of Yunnan by using relative data from 2000 to 2008 [7]. After researching, he believes the GDP of Yunnan province and the household income of residents will influence housing prices [7]. Chen used multiple regression analysis to investigate the variables that affect the costs of commercial housing in Changzhou City [8]. It is revealed that the average sale price of land has the greatest impact on house prices [8]. On this basis, this study gives some policy suggestions for some governments to formulate and regulate housing prices.

Using Grey correlation analysis, Yu and Jia investigated the factors that affect property prices in Tongling City and ranked them according to their significance [9]. Their findings indicate that GDP has the greatest impact on home prices [9]. However, in their report, they only included 5 aspects to be examined. The cost of real estate could be influenced by several factors. Wang et al. investigated the variables influencing Guizhou's housing costs [10]. The findings demonstrate that in Guizhou Province, GDP and per capita income have the greatest effects on house prices [10]. Based on this, the article gives some advice to policymakers from the supply and demand side.

Although some academics employ other research techniques to study this subject, multiple linear regression is the method of choice for most of them. According to Qing, compared with Spearman correlation coefficient method, the linear regression model is a good method for researching the price of real estate [11]. In fact, few academics are now using data from Yunnan and Guizhou provinces over the previous ten years to analyse and forecast the trend of house prices in these two areas. Therefore, this article will use multiple linear regression to research factors influencing the real estate price of Yunnan province and Guizhou Province. Then, offering policymakers and investors some advice

Yunnan province and Guizhou province are chosen to be the research objects. Yunnan Province and Guizhou Province are in southwest China and have similar landscape features. This article intends to use Yunnan province as the main research object and Guizhou province as the comparative test object. Yunnan province is one of China's most famous tourist provinces because its climate is warm all year. In recent years, more and more people go to Yunnan province to buy or rent houses to live there or as a kind of investment. Therefore, it is important to find some factors that would influence the price of real estate to give some suggestions to buyers and sellers. Guizhou province, which is close to Yunnan province, will be used as a comparative object to see whether these factors have similar impacts.

2. Influencing Variables of Housing Costs & Data

In this article, four different factors will be selected and investigated for the impact they have on the cost of housing. These four elements will be discussed in this section, followed by an explanation of the data sources and calculation techniques used for some of the data.

The first factor is the Loan Prime Rate (LPR), or long-term lending rate. LPR in China is the loan market quoted interest rate calculated and published by the Bank of China and 18 banks. The mortgage rate is added to the LPR. The customer's credit rating, in addition to other considerations, is considered when determining the mortgage rate. According to Zou's research, fluctuations in interest rates have a considerable impact on property values in China [12]. As a result, this essay chooses interest rates as one of its research variables. The second factor is the residential consumer price index. Housing expenditure is an important part of household expenditure. The residential price consumption index can reflect the consumption power of residents in terms of residence. Housing consumption power among inhabitants is likely to have an impact on housing demand, which will then have an impact on housing costs. The third factor is the amount of completed investment in real estate development. The completed investment amount of real estate development has a significant influence on the supply of real estate, and the price of housing is influenced by the supply of real estate. Zhan and Sun believe real estate values are significantly influenced by the amount of investment made in real estate development [13]. The finished commercial housing space makes up the fourth factor. Since housing is a commodity, supply and demand have an impact on its price. The housing supply is directly impacted by the amount of completed housing. The price of housing will decrease under the condition of constant demand if there is a high percentage of completed commercial housing units and vice versa.

This study primarily gathers and analyses monthly data on real estate in the provinces of Yunnan and Guizhou from 2013 to 2022. Specifically, the information primarily focuses on the average housing cost, LPR, or long-term loan interest rate, the residential consumer price index finished real

estate development investments, and the total amount of completed commercial housing area over the previous ten years. In this work, this article will first do a correlation analysis using these data, and then create a regression model to predict home prices using the model. The data was gathered from the National Bureau of Statistics of China, the Yunnan Provincial Bureau of Statistics, the Guizhou Provincial Bureau of Statistics, and the Oriental Fortune websites.

3. Variable Design & Correlation Analysis

In this section, firstly, the independent and dependent variables of this essay will be introduced. Then, this paper will conduct a correlation analysis on the selected independent variables and dependent variables. The Spearman correlation coefficient will be applied.

3.1. Variable Design

The dependent variables in this paper are the monthly average housing price data of Yunnan Province and Guizhou Province from 2013 to 2022, respectively. There are four independent variables in this study, as determined by the examination of factors impacting house prices in the second section. Regarding the classification of independent variables, this study analyses four housing price influencing factors based on the supply and demand of real estate. Among them, the Loan Prime Rate (LPR) or long-term lending rate and residential consumer price index are likely to affect investors' demand for real estate and then influence the housing price. Investment completed in real estate development and the total completed area of commercial housing will influence the availability of real estate and, consequently, the price of housing. Variable definitions are shown in Table 1.

Table 1. Variable definition

Types of variables	Variable name		Variable symbol
Independent variables	the Loan Prime Rate (LPR), or long-term lending rate		LA
	residential consumer price index		RCPI
	investment completed in real estate development	Yunnan province	IC_Y
		Guizhou province	IC_G
	total completed area of commercial housing	Yunnan province	CA_Y
		Guizhou province	CA_G
dependent variables	Average house price	Yunnan province	AHP_Y
		Guizhou province	AHP_G

3.2. Correlation Analysis

This study measures the correlation between variables using the Spearman correlation coefficient. Zhang's research indicates that the Spearman correlation coefficient has been widely employed in several correlation investigations [11]. Therefore, this paper chooses this method to study the correlation between variables. According to the data collected, this paper's correlation study includes five variables. The SPSS outcomes are displayed in Table 2. As given in Table 2, four independent variables derived from the data of Yunnan and Guizhou Province are significantly correlated with housing prices at the 0.01 and 0.05 significance levels. However, it can also be seen that there are also correlations among the independent variables. This study will use SPSS to run a multicollinearity test on independent variables in order to determine whether multicollinearity exists among

independent variables. It can be seen from Table 2 that four independent variables derived from the data of Yunnan and Guizhou Province are significantly correlated with housing prices at the 0.01 and 0.05 significance levels. However, it can also be seen that there are also correlations among the independent variables. This study will use SPSS to run a multicollinearity test on independent variables in order to determine whether multicollinearity exists among independent variables. According to Table 3, the VIF of all independent variables is less than 10. Thus, the multicollinearity among the independent variables has little effect on the regression model. The multicollinearity issue among the independent variables can be disregarded.

Table 2. Correlation analysis of variables in Yunnan and Guizhou Province

Pearson Correlations				
	LA	RCPI	IC	CA
AHP	-0.709**(Yunnan)	-0.504**(Yunnan)	0.666**(Yunnan)	-0.224*(Yunnan)
	-0.722**(Guizhou)	-0.304**(Guizhou)	0.503**(Guizhou)	-0.426**(Guizhou)
**. Correlation is significant at the 0.01 level (2-tailed)				
*. Correlation is significant at the 0.05 level (2-tailed)				

Table 3. Results of multicollinearity test of Yunnan and Guizhou province

Coefficients		
Model	Collinearity Statistics	
	VIF (Yunnan)	VIF (Guizhou)
LPR/Long-term loan interest rate	2.696	1.698
Residential consumer price index	2.597	1.581
Investment completed in real estate development(thousand)	1.213	1.421
Total completed area of commercial housing(thousand)	1.011	1.403

4. Multiple Linear Regression

4.1. Model Establishment

Real estate values are influenced by several factors. Cao et al. stated that multiple linear regression models are appropriate for investigating economic problems in which a single variable is influenced by multiple variables [14]. As a result, this paper employs a multiple regression model for modelling. The Yunnan Province housing pricing model is shown below (Model 1).

$$AHP_Y = \beta_0 + \beta_1 LA + \beta_2 RCPI + \beta_3 IC_Y + \beta_4 CA_Y + \varepsilon. \quad (1)$$

In model 1, β_0 is a constant term. β_1, β_2, β_3, β_4 are the regression coefficients. ε is an error term. The housing price model of Guizhou Province (Model 2) is shown as:

$$AHP_G = \alpha_0 + \alpha_1 LA + \alpha_2 RCPI + \alpha_3 IC_G + \alpha_4 CA_G + \varepsilon \quad (2)$$

In model 2, α_0 is a constant term. α_1, α_2, α_3, α_4 are the regression coefficients, ε is an error term.

4.2. Analysis of Model Regression Results

The collected data are analysed using multiple linear regression with the aid of SPSS. The regression coefficient results for Yunnan Province are presented in Table 4. According to the regression results (Table 4), the Sig values of the three independent variables LA, ICY and CAY are all less than 0.05. Therefore, the independent variable is significant in the model. However, the Sig. value of RCPI is equal to 0.05, and this independent variable is not significant in the model. Consequently, model modification is required. The regression coefficient results of Guizhou Province are shown in Table 5. The sig values of the independent variables LA and ICG are less than 0.05, whereas the sig values of RCPI and CAG are greater than 0.05, as shown in Table 5. Consequently,

it is inappropriate to incorporate four independent variables into a single model. Subsequently, this paper will modify the housing price model of Yunnan and Guizhou provinces.

Table 4. Results of regression coefficient of Yunnan Province

Model	Unstandardized Coefficients	Standardized Coefficients		t	Sig.
	B	Std. Error	Beta		
Constant	-7964.398	9021.221		-0.883	
LA	-1218.001	158.917	-0.653	-7.664	9.0295E-12
RCPI	187.759	94.559	0.166	1.986	0.05
IC_Y	0.000069	0.000009	0.448	7.843	3.6735E-12
CA_Y	-0.041	0.013	-0.168	-3.217	0.002
Dependent Variable: Average house price					

Table 5. Results of regression coefficient of Guizhou Province

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	1617.750	4351.107		-0.372	
LA	-583.850	88.815	-0.575	-6.574	1.8932E-9
RCPI	56.971	44.055	0.109	1.293	0.199
IC_G	0.000035	0.000009	0.312	3.903	0.000167
CA_G	-0.007	0.01	-0.057	-0.717	0.475
Dependent Variable: Average house price					

4.3. Regression Model Modification

According to Cheng and Jia, the relationship between the independent variable and the dependent variable can be studied using a model with a single independent variable and a dependent variable [15]. This paper uses the independent variable RCPI and the dependent variable Average house price to establish a regression model for the model modification of Yunnan Province and obtains Model 3. Model 4 is obtained by putting the independent variables LA, CAY and ICY into a regression model with the dependent variable Average house price. Model 3 is shown as follows.

$$AHP_Y = \beta_0 + \beta_5 RCPI + \varepsilon \tag{3}$$

In Model 3, β_0 is a constant term. β_5 is the regression coefficient. ε is an error term. Model 4 is shown below.

$$AHP_Y = \beta_0 + \beta_6 LA + \beta_7 IC_Y + \beta_8 CA_Y \tag{4}$$

In Model 4, β_0 is a constant term, β_6, β_7, β_8 are the regression coefficients. ε is an error term. The regression coefficients of Model 4 are obtained by using SPSS in the following Table 6.

RCPI is significant in the regression model, according to the results of the regression. It is evident that a rise in the independent variable RCPI will result in a fall in the dependent variable Average house price. The regression coefficients of Model 4 are obtained by using SPSS in the following Table 7. According to the regression equation in Table 7, the independent variables LA and CAY have a negative relationship with average house prices. In addition, there is a positive correlation between ICY and the average house price.

Table 6. Results of regression coefficient of model 3

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	64110.854	9483.735		6.760	7.0933E-10
RCPI	-570.040	93.607	-0.504	-6.089	1.7293E-8
Dependent Variable: Average house price					

Table 7. Results of regression coefficient of model 4

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Constant	9891.838	726.016		13.625	4.0395E-25
LA	-982.667	107.313	-0.527	-9.157	4.1187E-15
IC_Y	0.000067	0.000009	0.439	7.602	1.1853E-11
CA_Y	-0.043	0.013	-0.174	-3.217	0.001
Dependent Variable: Average house price					

For model modification of Guizhou Province, this paper will establish regression models with independent variables RCPI and CAG and dependent variable Average house price respectively, to obtain Model 5 and Model 6. At the same time, the rest independent variables will be put into a regression model to obtain Model 7. Model 5 is shown as follows.

$$AHP_G = \alpha_0 + \alpha_5 RCPI + \varepsilon \tag{5}$$

In this model, α_0 is a constant term. α_5 is the regression coefficient. ε is an error term. Model 6 is shown below.

$$AHP_G = \alpha_0 + \alpha_6 CA_G + \varepsilon \tag{6}$$

In Model 6, ∂_0 is a constant term. ∂_6 is a regression coefficient. ε is an error term. Model 7 is shown as follows.

$$AHP_G = \alpha_0 + \alpha_7 LA + \alpha_8 IC_G + \varepsilon \tag{7}$$

In Model 7, α_0 is a constant term. α_7, α_8 are regression coefficients. ε is an error term. The regression coefficients of Model 5 are obtained by using SPSS in the following Table 8.

Table 8. Results of regression coefficient of model 5

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	22868.055	4757.373		4.807	0.000005
RCPI	-175.894	44.055	-0.109	-3.737	0.000298
Dependent Variable: Average house price					

Table 9. Results of regression coefficient of model 6

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	5472.818	102.810		53.232	7.6468E-80
CA_G	-0.054	0.011	-0.419	-4.819	0.000005
Dependent Variable: Average house price					

Table 10. Results of regression coefficient of model 7

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	7233.773	486.091		14.882	6.4887E-28
LA	-558.345	73.620	-0.550	-7.584	1.2405E-11
IC_G	0.000033	0.000008	0.296	4.078	0.000087
Dependent Variable: Average house price					

RCPI is significant in the regression model (Table 8). The average housing price in Guizhou Province will increase with the decline of the residential consumer price index. A negative relationship exists between them. The regression coefficients of Model 6 are obtained by using SPSS in the following Table 9. Therefore, the regression coefficient between the average housing price and the total completed area of commercial housing in Guizhou Province is negative. The regression coefficients of Model 7 are obtained by using SPSS in the following Table 10. The independent variables in the model are significant. The regression coefficients show negative relationship between

the LPR or long-term lending rate and the dependent variable. In contrast, the average house price will fall as real estate development investment increases.

4.4. Regression Model Testing

The modified model (Model 3-7) was tested according to SPSS. The results of the model significance test are summarized in the following Table 11. Table 11 shows that the Sig value of the modified model is less than 0.05. Therefore, the adjusted regression model is statistically significant.

Table 11. Results of model significance test

	Model 3	Model 4	Model 5	Model 6	Model 7
F	37.08	84.962	13.966	23.219	56.293
Sig.	1.7293E-8	3.4164E-28	0.000298	0.000005	1.7855E-17
adjusted R^2	0.247	0.696	0.105	0.168	0.501

5. Conclusion

In summary, in terms of method, this paper uses SPSS and multiple linear regression models to study the monthly housing price data of Yunnan Province and Guizhou Province over the past 10 years. In terms of data, this paper collects monthly housing price data of Yunnan and Guizhou provinces and the data of LPR or long-term loan rate, residential consumer price index, investment completed in real estate development and the total completed area of commercial housing from 2013 to 2022. This research paper reaches the four following conclusions. Firstly, housing prices in Yunnan and Guizhou provinces will rise with the decline of long-term loan interest rates. The regression coefficient indicates that, compared to Guizhou province, the change in the long-term loan interest rate in Yunnan province exerts a greater influence in the opposite direction of the housing price in Yunnan province. Secondly, the residential consumer price index has a negative relationship with housing prices in Yunnan and Guizhou provinces. Under the condition that residents' income is stable, if their consumption of other things increases, it is likely that they do not have enough money to buy houses. This will result in a decline in the demand for real estate among residents, which will lead to a decline in real estate prices. Thirdly, an increase in the amount of real estate investment will lead to an increase in housing prices. For real estate consumers or investors, an increase in the amount of money invested in real estate is likely to give them a positive signal that the real estate market will become better. Based on the predicted impact of investors, they are likely to have an impulse to buy houses, which in turn will increase the demand for houses and raise prices. Fourthly, the completed area of real estate has an inverse relationship with house prices. The Total completed area of commercial housing increased, meaning the house supply increased. With stable demand, the price of a house will be lower. This paper makes a comparative study between the data of Yunnan Province and Guizhou Province. It is found that the above four conclusions are applicable to the real estate market in Yunnan Province as well as Guizhou Province.

Based on the above conclusions, this paper will give some suggestions to policymakers, real estate enterprises, housing investors or consumers. Firstly, policymakers can adjust housing prices by adjusting long-term loan rates in the market. If they do, they may be able to influence consumer demand for real estate through the policy. Secondly, real estate enterprises could refer to the residential consumer price index and the total completed area of commercial housing before deciding the future development of real estate projects. If the residential price index or total completed area of commercial housing is high, housing prices may decline. Real estate enterprises can invest prudently according to these signals before investing. Thirdly, it is important for real estate consumers to consider their current financial conditions and the real estate market before deciding to buy real estate. Especially as investment completed in real estate increases, investors need to think rationally about investment decisions.

References

[1] Alfiyatin, A. N., Ruth, E. F., Hilman T., Mahmudy, W. F.: Modelling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study: Malang, East Java, Indonesia. International Journal of Advanced Computer Science & Applications, 8 (10), (2017).

[2] Zhou, L., Xia, E.: A Review of Factors Influencing Housing Prices in Foreign Countries. Technoeconomics, 37 (12), 111-119 (2018).

[3] Li, X.: Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Model. Computational Intelligence and Neuroscience, 2022, 1–9 (2022).

[4] Liu, G.: Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model. Scientific Programming, (2022).

[5] Chen, N.: House Price Prediction Model of Zhaoqing City Based on Correlation Analysis and Multiple Linear Regression Analysis. Wireless Communications and Mobile Computing, 2022, 1–18 (2022).

[6] Koohpayma, J., Argany, M.: Estimating the price of apartments in Tehran using extracted compound variables. International Journal of Housing Markets and Analysis, 14 (3), 569–595 (2021).

[7] Jian, X.: Empirical Analysis of Influencing Factors of Commercial Housing Prices in Yunnan Province. Cooperative Economy and Science & Technology, 654, 62-63 (2021).

[8] Chen, L.: Analysis of Influencing Factors of Commercial Housing Prices in Changzhou City. Real Estate World, 379, 25-28 (2022).

[9] Yu, B., Sheng, J.: Analysis of Factors Influencing Housing Prices in Tongling City. Cooperative Economy and Science & Technology, 668, 65-67 (2021).

[10] Wang, L., Liu, P., Wang, M., et al.: Analysis of Factors Influencing Housing Prices in Guizhou Province Based on VAR Model. Computer Knowledge and Technology, 18 (02), 122-125+129 (2022).

[11] Zhang, Q.: Housing Price Prediction Based on Multiple Linear Regression. Scientific Programming, (2021).

[12] Zou, X.: Research on the Effect of Interest Rate Adjustment on Asset Prices in China: An Analysis Based on Feature Facts and VAR Model. China Price, 385, 47-51 (2022).

[13] Zhan, S., Sun, C.: Analysis of Influencing Factors of Housing Prices in Shenyang City Based on VAR Model. Journal of Shenyang Jianzhu University (Social Science Edition), 24 (02), 167-172 (2022).

[14] Cao, Q., Zhu, J., Jin, J.: Econometric Analysis of Factors Affecting Tax Revenue in China Based on Multiple Linear Regression. Journal of Harbin Normal University (Natural Science Edition), 36 (04), 9-15 (2020).

[15] Chen, C., Jia, M.: The Relationship between Equity Structure and Operating Performance under the Background of Mixed Ownership Reform: An Empirical Analysis Based on Listed Banks in China. China Business Review, 2019 (10), 153-155 (2019).