ISSN: 2688-9323

Analysis of Factors Influencing the Average Price of Commodity House Sales in Anhui Province

Huaqin Shi

School of International Trade and Economics, Anhui University of Finance and Economics, Bengbu 233000, China.

Abstract

Based on the continuous development of China's economy in recent years and the rising per capita income of our residents, this paper uses econometric knowledge for the cross-sectional data of 16 famous cities in Anhui Province in 2020 to explore and establish an econometric model on the influencing factors of house prices to determine whether per capita disposable income, public transportation and green space have an impact on the sales price of commercial houses and to conduct research.

Keywords

House Price; Disposable Income Per Capita; Public Transportation; Green Space.

1. Introduction

Since the Third Plenary Session of the Eleventh Central Committee, China has implemented the policy of reform and opening up, and accordingly started the exploration of housing system reform in the real estate market, which has led to the rapid development of China's real estate industry, especially after China has established the socialist market economy system, the real estate industry has developed more rapidly. In 1998, the State Council issued the Notice on Further Deepening the Reform of Urban Housing System and Accelerating Housing Construction, which stopped the distribution of housing in kind and gradually implemented the monetization of housing distribution, so that China's urban housing system gradually reformed from the planned to the market economy style and began to form a market-oriented housing system.[1] The market-based housing system began to take shape. Since then, some major cities in China, such as Beijing, Shanghai and Guangzhou, have continued to heat up the real estate industry and real estate prices have risen steadily [2].

With the development of China's economy, in recent years, the overheated real estate market and the rapid rise in housing prices have triggered a series of negative effects affecting social stability, bringing different degrees of impact to China's economic development and social residents, and becoming a hot topic of general concern to society and the public. Real estate as a special commodity, its price is affected by the national macroeconomic policy, the market, economic conditions and other factors, due to the rapid rise in housing prices and the lack of other better investment channels in the market, real estate is increasingly becoming the main direction of people's investment and consumption.

It is of great importance to scientifically understand the development law of real estate market and objectively evaluate the current situation of real estate market development for residents' investment and consumption planning as well as to provide basis for government departments' decision making. Under such circumstances, it is quite important to predict the house price accurately. This paper focuses on the cross-sectional data of 16 cities in Anhui Province in 2020 to explore the relationship between their house price data and their influencing factor data.

2. Literature Review and Research Hypothesis

The real estate industry in foreign countries has a long history and has gradually formed a perfect market mechanism. Therefore, there is also more research literature on this aspect abroad, and foreign countries attach more importance to quantitative analysis and practical application, and there is also a lot of literature on its empirical analysis.

Miller (2006) used data from 277 US regions to establish a VAR model to analyze the relationship between dynamic real estate price fluctuations and changes in per capita income growth rate, population growth rate, unemployment rate, and per capita GDP, and the results showed that there is a correlation between real estate price volatility and per capita income growth rate, future real estate prices, and per capita GDP.

The domestic real estate industry is relatively late to be marketed, and scholars' research in this area is also relatively late compared to the West. However, with the rapid rise of the industry in China in the past few years, real estate prices have also been rising, and this phenomenon has begun to attract the attention of domestic scholars and experts, and more people have begun to devote themselves to the study of real estate prices. Ma Zhefeng and Yuan Hui (2008) established a VAR model through a total of 38 quarters of data from 1999 to 2008 in China, and after empirical analysis, they concluded that land prices, money supply and economic growth are all positively related to real estate prices.

From the existing research, most of the perspectives are for the influence of variables such as income, cost and GDP on the house price. And with the continuous development of China's economy, the rise of housing prices in various regions has been rising, from the other hand, it can also be seen that people's purchasing power has risen, but in nowadays people do not only decide the level of housing prices based on income, as people pursue more and more, some factors of urban development also need to be taken into account. For example, the convenience of transportation, we know that in places with convenient transportation, its commodity housing prices must not be low, it can affect people's travel; In addition, now everyone is beginning to pursue a fresh environment, so the environment around the housing will also have an impact on housing prices. According to the above theoretical analysis, the direction of this paper's research hypothesis is proposed: this paper focuses on three aspects of external economic variables - per capita disposable income, traffic condition and green area of the area - to study whether there is an influence on the sales price.

3. Model Setting and Variable Description

3.1. Variable Description

3.1.1. Analysis of Influencing Factors

Disposable income per capita. Disposable income per capita refers to the total amount of personal discretionary savings and consumption, which reflects the living standard and purchasing power of people in a country or region. Generally speaking, housing demand increases with the increase of disposable income per capita. Considering that the average price of selling commercial housing is being explored here, disposable income per capita is cited as one of the explanatory variables.

Public transportation. It is well known that the accessibility of the transportation nowadays has become a major factor in the price of housing. Housing requires consideration of many aspects, and accessibility is particularly important for commuters and school-goers. Therefore, the state of public transportation is used as one of the explanatory variables affecting house prices.

Green area. Green area is considered because it is necessary to take into account the green ratio when buying a house, which reflects the environment of the house from the side. In today's society where people are more and more concerned about the environment, green area has

ISSN: 2688-9323

become an aspect that affects the price of housing. The green area and green ratio are proportional, the larger the green area, the higher the green ratio. Considering the difficulty of collecting green ratio data around the world, this paper takes the green area as another explanatory variable that affects the house price.

Other factors. Of course, the influence of house price is not only the above factors, in this paper other factors as a random factor, collectively classified as a random disturbance term ε_i.

3.1.2. Data Collection

In this paper, we collected the data shown in Table 1 through the Statistical Yearbook of Anhui Province 2021.

Table 1. Sample data

	Average price of commercial housing (ten thousand / m²)	Per capita disposable income (RMB)	Public transport (vehicle)	Green space area (ha)
Hefei City	14321	41619.2	8370	20271
Huaibei City	6554	28126.6	767	4781
Bozhou city	5960	22274.3	1805	4085
the City of Suzhou	5648	22105.3	1368	3340
Bengbu City	6149	29247.2	1888	6038
Fuyang city	6667	22238.6	1818	6814
Huainan City	6361	28780.2	1203	5193
Chuzhou	6211	25711.4	1772	5293
Lu'an City	6405	22457.4	2197	3466
Ma'anshan City	7726	42392.4	853	6355
Wuhu City	8756	36828.6	1860	9870
Xuancheng City	6264	30745.9	1353	4259
Tongling city	5539	29567.9	966	6981
Chizhou city	6353	26404.3	484	1857
Anqing City	6714	24647	1407	6199
Huangshan city	7257	27916.4	220	13498

3.2. Model Setting

Due to the difficulty of data collection, this paper uses sample data from 16 cities in Anhui Province in 2020 and focuses on the effects of per capita disposable income, urban transportation, and green area on the average price of commercial housing sales.[3] The average price of commodity houses is taken as the dependent variable, and the per capita disposable income, public transportation and green space area are taken as the independent variables. Using econometric theory, a multiple linear regression model is constructed to explore the relationship between its house price data and its influencing factor data. The linear regression model is as follows:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + e_i \quad (1)$$

ISSN: 2688-9323

Table 2. variable description

dependent variable Y	Average price of commercial housing (ten thousand / m^2)
argument X	Per capita disposable income (RMB)
	Public transport (vehicle)
	Green space area (ha)

4. Model Estimation

Parameter estimation was performed on the selected data using ordinary least squares (OLS) method with respect to EViews 9.0 software, and the results are shown in Figure 1.

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 12/17/22 Time: 09:48
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2390.129	723.8061	3.302167	0.0063
X1	0.089323	0.029169	3.062264	0.0099
X2	0.585244	0.116224	5.035483	0.0003
X3	0.155854	0.053379	2.919774	0.0128

R-squared	0.938161	Mean dependent var	7055.313
Adjusted R-squared	0.922701	S.D. dependent var	2093.319
S.E. of regression	581.9996	Akaike info criterion	15.78313
Sum squared resid	4064683.	Schwarz criterion	15.97628
Log likelihood	-122.2651	Hannan-Quinn criter.	15.79303
F-statistic	60.68377	Durbin-Watson stat	2.312915
Prob(F-statistic)	0.000000		

Figure 1. regression results

From the least squares method, the preliminary results of the regression model are as follows:

$$Y = 2390.129 + 0.0893X_1 + 0.5852X_2 + 0.1559X_3 \quad (2)$$

5. Model Testing

After the initial parameter estimation, we cannot directly apply the model to the future forecasts. Model testing must be performed, and only after passing the test can the model be used. Because this paper uses the ordinary least squares (OLS) method for parameter estimation, the use of least squares requires that the preconditions are met before the model can be used. For this reason, we need to perform three tests, namely economic test, statistical inference test, and econometric test, before we can obtain the final model to be used in practice.

5.1. Economic Test

According to the model we can see that for every dollar increase in per capita disposable income, the average sales price of commercial housing will increase by 0.0893 million/m2; for every increase in public transportation, the average sales price of commercial housing will increase by 0.5852 million/m2; for every hectare increase in green space, the average sales price of commercial housing will increase by 0.1559 million/m2. And the model results are consistent with the economic theory according to the reality, so it passes the economic test successfully.

ISSN: 2688-9323

5.2. Statistical Inference Test

5.2.1. Goodness-of-fit Test

R2=0.9382>0.8, which means that 93.82% of the changes in the explanatory variable Y in the model are caused by the changes in the explanatory variable X in the model. Since the decidable coefficient R2 is greater than 0.8, the model successfully passes the goodness-of-fit test.

5.2.2. Significance Test of Variables (T-test)

From Figure 1, we can see that the t-tests of β1, β2, and β3 have t-values of 3.0623, 5.0355, and 2.9198, respectively. at the significance level $\alpha = 0.05$, since the probabilities are all less than 0.05, and the critical value of 2.1788 is obtained by checking the table. obviously, the t-values of each explanatory variable are greater than the critical value, so the original hypothesis is rejected, and X1, X2, and X3 are significant and smoothly passed the t-test.

5.2.3. The Significance Test of the Equation (F-test)

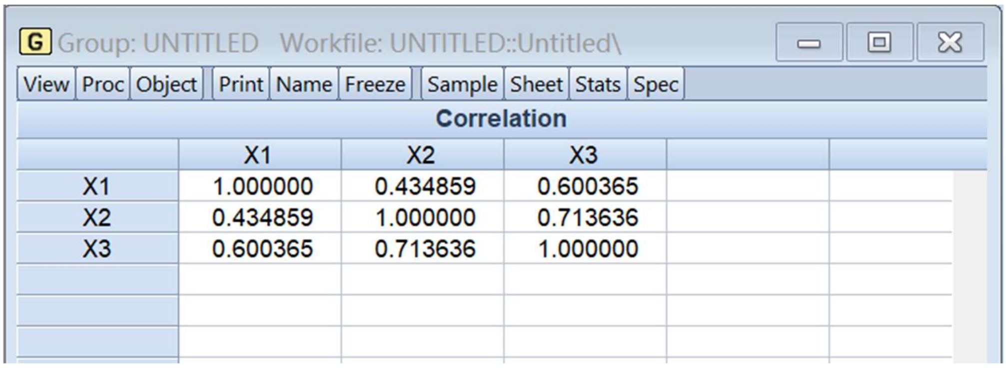
From the ordinary least squares (OLS method), it is known that F=60.68377, at the given significance level α=0.05, and since the Prob of F test is 0.000000, check the F distribution table to find F(3,12)=3.49. Obviously, the F value is greater than the critical value, so the original hypothesis is rejected, and the joint effect of all explanatory variables in the model on the explained variables is significant. The model successfully passed the F-test.

5.3. Econometric Tests

5.3.1. Multicollinearity Test

To confirm whether there is multicollinearity in the model, i.e., whether there is a strong linear correlation between the explanatory variables, this paper adopts two ways to test.

Simple correlation coefficient method.Enter the command: COR X1 X2 X3 in EViews to obtain the correlation coefficient matrix. If the correlation coefficient of each explanatory variable is high (greater than 0.8), there must be multicollinearity; otherwise, there is not. As shown in Figure 4, the correlation coefficients among the explanatory variables are not higher than 0.8, so we judge from the preliminary judgment that there is no multicollinearity in the model.


Group: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Correlation

	X1	X2	X3
X1	1.000000	0.434859	0.600365
X2	0.434859	1.000000	0.713636
X3	0.600365	0.713636	1.000000

Figure 2. correlation coefficient test results

Variance-inflation factor method.To further understand the nature of multicollinearity, we did an auxiliary regression by regressing each X variable separately as an explanatory variable for the X variables under it, and calculating the decidable coefficients and variance inflation factors. If the VIF (1/(1-R2)) corresponding to each explanatory variable did not exceed 10, there was no multicollinearity. If one VIF is greater than 10, the model has multicollinearity and must be corrected. The operation is performed by EViews software and the results are shown in Table 3, where we find that there is no multicollinearity.

ISSN: 2688-9323

Table 3. VIF Results

dependent variable	R^2	VIF
X1	0.3605	1.5637
X2	0.5093	2.0379
X3	0.613	2.584

In addition, we also used a direct calculation in EViews to obtain the value of VIF as in Figure 3. Therefore, the model passed the test successfully and the equations were free from multicollinearity.

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Variance Inflation Factors
Date: 12/16/22 Time: 19:24
Sample: 1 16
Included observations: 16

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	523895.2	24.74680	NA
X1	0.000851	34.93660	1.563776
X2	0.013508	4.038625	2.038073
X3	0.002849	8.750461	2.584065

Figure 3. Variance inflation factor results

5.3.2. Heteroskedasticity Test

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: White

F-statistic	0.733499	Prob. F(9,6)	0.6758
Obs*R-squared	8.381851	Prob. Chi-Square(9)	0.4962
Scaled explained SS	5.940274	Prob. Chi-Square(9)	0.7459

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/16/22 Time: 19:35
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-14845550	10314164	-1.439336	0.2001
X1^2	-0.014559	0.008107	-1.795902	0.1226
X1*X2	-0.205243	0.161282	-1.272569	0.2503
X1*X3	0.023143	0.023253	0.995281	0.3580
X1	1009.946	619.7240	1.629670	0.1543
X2^2	-0.071688	0.261161	-0.274496	0.7929
X2*X3	0.193326	0.163979	1.178965	0.2830
X2	4511.061	3998.217	1.128268	0.3023
X3^2	0.007300	0.030268	0.241191	0.8174
X3	-913.0155	723.7105	-1.261576	0.2539

R-squared	0.523866	Mean dependent var	254042.7
Adjusted R-squared	-0.190336	S.D. dependent var	416493.3
S.E. of regression	454404.7	Akaike info criterion	29.16053
Sum squared resid	1.24E+12	Schwarz criterion	29.64340
Log likelihood	-223.2843	Hannan-Quinn criter.	29.18526
F-statistic	0.733499	Durbin-Watson stat	1.654939
Prob(F-statistic)	0.675761		

Figure 4. White Test Results

Since the existence of heteroskedasticity will have a series of negative effects on the regression model, we adopt two tests to determine whether there is heteroskedasticity in the model.

White test method.The results of the test are as follows (Figure 6). At the significance level α=0.05, Obs*R-squared(nR2)=8.3819, nR2 asymptotically obeys the χ^2 distribution, and at the given significance level, checking the χ^2 distribution table yields its critical value $\boldsymbol{\chi^2_{0.05}(9)=16.919}$, nR2<$\boldsymbol{\chi^2_{0.05}(9)}$ and its Prob is 0.4962>0.05, so the original hypothesis is accepted and the model is considered to have no heteroskedasticity.

Park test.Heteroskedasticity was tested by constructing an auxiliary regression model:

$$\mathrm{ln}e_i^2=\mathrm{ln}\alpha+\beta_1\mathrm{ln}X_1+\beta_2\mathrm{ln}X_2+\beta_3\mathrm{ln}X_3+v_i. \quad (3)$$

The regression model:

$$\mathrm{ln}e_i^2=1.0493+1.6651\mathrm{ln}X_1-0.2170\mathrm{ln}X_2-0.6424\mathrm{ln}X_3 \quad (4)$$

F=0.1604, Prob=0.9209>0.05, so the original hypothesis is accepted and the model is considered to have no heteroskedasticity.

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E
Method: Least Squares
Date: 12/16/22 Time: 20:02
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.049329	28.85748	0.036362	0.9716
X11	1.665064	3.259335	0.510860	0.6187
X22	-0.216977	0.749151	-0.289631	0.7770
X33	-0.642420	1.235878	-0.519808	0.6127

R-squared	0.038553	Mean dependent var	10.99028
Adjusted R-squared	-0.201809	S.D. dependent var	1.974711
S.E. of regression	2.164817	Akaike info criterion	4.594867
Sum squared resid	56.23720	Schwarz criterion	4.788014
Log likelihood	-32.75893	Hannan-Quinn criter.	4.604758
F-statistic	0.160395	Durbin-Watson stat	2.085858
Prob(F-statistic)	0.920924		

Figure 5. Park's test results

5.3.3. Autocorrelation Test

The absence of correlation between random perturbation terms corresponding to different sample points is one of the classical assumptions of linear regression models, i.e., there is no autocorrelation, and to verify this assumption we test whether the model has autocorrelation by three tests.

DW test method.DW=2.3129, and the DW statistics table shows that dL=0.857 and du=1.728. 4-dU<DW<4-dL in the model, so it cannot be determined whether there is autocorrelation.

Partial correlation coefficient test method.Since the DW test is strict and cannot test for autocorrelation, the partial correlation coefficient test has wider applicability, so we use EViews software to test it. We can see that the column of Partial Correlation is always within the dashed line, i.e., there is no autocorrelation.

ISSN: 2688-9323

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals

Date: 12/16/22 Time: 20:39
Sample: 1 16
Included observations: 16

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.162	-0.162	0.5031	0.478
		2	-0.236	-0.270	1.6522	0.438
		3	-0.047	-0.154	1.7013	0.637
		4	-0.044	-0.172	1.7477	0.782
		5	0.177	0.087	2.5719	0.766
		6	-0.197	-0.230	3.6949	0.718
		7	-0.022	-0.067	3.7104	0.812
		8	0.279	0.200	6.5046	0.591
		9	-0.146	-0.095	7.3836	0.597
		10	-0.065	-0.053	7.5894	0.669
		11	-0.204	-0.249	9.9953	0.531
		12	0.062	-0.098	10.270	0.592

Figure 6. Partial correlation coefficient test result

B-G test method.The results of the test are as follows, we can see that Obs*R-squared is the nR2 statistic with a value of 2.376456, and the Prob=0.3048>0.05 for nR2 at the significance level α=0.05, which is greater than the significance level, so we can assume that the model does not have autocorrelation.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.872187	Prob. F(2,10)	0.4476
Obs*R-squared	2.376456	Prob. Chi-Square(2)	0.3048

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 12/16/22 Time: 20:53
Sample: 1 16
Included observations: 16
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	34.71764	741.4091	0.046827	0.9636
X1	-0.008409	0.030634	-0.274507	0.7893
X2	-0.075465	0.130755	-0.577149	0.5766
X3	0.048159	0.065446	0.735849	0.4787
RESID(-1)	-0.291550	0.317822	-0.917337	0.3806
RESID(-2)	-0.428760	0.360434	-1.189567	0.2617

R-squared	0.148529	Mean dependent var	-2.27E-13
Adjusted R-squared	-0.277207	S.D. dependent var	520.5563
S.E. of regression	588.2994	Akaike info criterion	15.87235
Sum squared resid	3460962.	Schwarz criterion	16.16207
Log likelihood	-120.9788	Hannan-Quinn criter.	15.88718
F-statistic	0.348875	Durbin-Watson stat	2.165781
Prob(F-statistic)	0.871664		

Figure 7. B-G test results

6. Conclusion

Based on the continuous economic development of China in recent years and the rising per capita income of China's residents, real estate data from the China Statistical Yearbook data in 2021 are matched with the per capita income of residents data and national economic

ISSN: 2688-9323

accounting data, and a multiple linear regression model is established using ordinary least squares method and using EViews software to collect the 2020 China Anhui Province The cross-sectional data of 16 cities in Anhui Province in 2020 were collected to investigate the impact of external economic variables: disposable income per capita, urban construction level: public transportation, and urban development level: green area on sales price. The results of the study show that disposable income per capita, public transportation and green space have positive effects on the sales of commercial properties. A multiple linear regression equation was obtained:

$$Y = 2390.129 + 0.0893X_1 + 0.5852X_2 + 0.1559X_3 \tag{5}$$

However, from a different perspective, we find that public transportation has more influence on the sales price of commercial properties in this study, and contrary to our hypothesis, the influence of disposable income per capita on housing prices is not significant, and it can be said that the fluctuation of housing prices in large cities in Anhui Province is relatively flat. We also found that more and more people are now considering the accessibility and living environment as the basis for the increase or decrease of house prices compared to the per capita disposable income. Therefore, we can see that the growth of property prices is increasingly related to the development of cities and environmental changes. For this reason, real estate developers need to take into account the impact of transportation and the environment when choosing a site, especially since access to public transportation has a certain degree of influence on whether people are willing to buy.

References

[1] Hui Zhou. China’s Monetary Policy Regulation and Financial Risk Prevention[M].Springer, Berlin, Heidelberg:2014.

[2] Information on:http://www.npc.gov.cn.

[3] Chen K . China's Real Estate Control Policies and Housing Price Fluctuations[C]// Proceedings of the 1st International Conference on Business, Economics, Management Science (BEMS 2019). 2019.