

Analysis on the Influencing Factors of Housing Prices in Shanxi Province Based on multiple Linear Regression

Jiamei Cheng*

College of Silesian Intelligent Science and Engineering, Yanshan University, Qinhuangdao, 066000, China

*Corresponding author: jm123@stumail.ysu.edu.cn

Abstract. As a pillar industry of the national economy, real estate price fluctuations affect residents' living quality, macroeconomic stability and urban development. Housing prices in Shanxi Province have also fluctuated in recent years, affecting people's living standards. However, there is no clear formula to reveal the essential reasons for changes in housing prices. To explore the law of housing price changes in Shanxi Province, this paper uses SPSS to build a multiple linear regression model, and analyzes the impact mechanism of various factors on housing prices from the perspective of supply and demand based on relevant data of Shanxi Province from 2013 to 2024. With regional GDP, per capita disposable income of urban residents, residential sales area, population, residential completion investment and residential completed area as independent variables, and the average sales price of commercial housing as the dependent variable, the study shows that residential commodity investment, regional GDP and residential commercial housing completion area have a great impact on housing prices in Shanxi Province, and the real estate industry is highly correlated with economic development and residents' income. Relevant suggestions are put forward to promote the healthy and steady development of Shanxi's real estate industry and keep housing prices in a reasonable range.

Keywords: Multiple linear regression; Housing price; Influencing factors; Optimization suggestions.

1. Introduction

In recent years, China's economy has maintained a steady growth momentum, but the irrational surge in housing prices has gradually become a prominent threat to both the economy and society. Zhou points out that the real estate sector has long been a pillar of China's economic growth, yet the exorbitant price increases before the pandemic have led to significant risks, such as housing prices being seriously out of line with household incomes and property developers accumulating heavy debts due to expectations of continued price rises [1].

Fluctuations in real estate prices have far-reaching impacts, affecting residents' lives, macroeconomics, and urban development. Dang highlights that the real estate and construction industry accounts for about 15% of China's GDP, so a downturn in the housing market after a long period of expansion could have a substantial adverse impact on the economy [2]. Wu conducts in-depth research on urban housing price influencing factors using SPSS, providing valuable methodological insights for related studies [3]. Wan focuses on the influencing factors of second-hand housing prices in Chongqing through a hierarchical model analysis, drawing academic attention to the variations in housing prices across different cities [4]. Xie carries out an econometric analysis of housing price influencing factors in Anhui using multiple regression, revealing that population changes are closely linked to housing demand [5]. Zhu explores statistical inference methods for housing price influencing factors, offering new perspectives for related research [6]. Zhang emphasizes that housing prices are fundamentally determined by the interaction of supply and demand [7]. Chu shows that rising housing prices will drive up retail price inflation, thereby increasing the living costs of households [8]. Qin Yanjiao's paper constructed multiple linear regression and random forest prediction models by conducting feature analysis, preprocessing, and partitioning on the public dataset of California's 2021 housing prices. After evaluating with MAE, RMSE, and R^2 , it concluded that the random forest model has better prediction accuracy [9]. Zhang Linhu's paper, using data of Dongying City from 2014 to 2020, built a multiple linear regression

model with SPSS, found that per capita GDP, urban population, and real estate investment have significant positive impacts on commercial housing prices, and put forward reasonable suggestions for housing price fluctuations [10].

Shanxi is currently facing the problem of rising housing prices with the development of its local economy. Scholars have categorized the factors affecting housing prices into five types. Dang 's study on Tangshan City finds that GDP is an important driver of housing price growth [2]; Xie and Zhu 's research (in Anhui and Dongying) indicates that population factors have a direct impact on housing demand [5, 6]. Zhou's study on Shanghai's tax policies related to the real estate market highlights the important role of policy factors [1]. Following Zhang 's supply-demand framework [7], this paper will use SPSS for regression analysis (referencing Wu 's methodology) to explore the factors driving housing price changes in Shanxi and put forward targeted suggestions [3].

2. Analysis of Housing Price Factors and Establishment Model

Table 1. Statistical Table of Real Estate-Related Data in Shanxi Province from 2013 to 2024

Variable	Sample size	Minimum value	Maximum value	Mean value	Standard deviation	Median
Gross Regional Product	12	11836.390	26051.000	17758.701	5758.985	16459.870
Per Capita Disposable Income	12	22455.600	43036.000	32437.991	6908.358	32148.700
Population	12	3445.960	3535.000	3497.431	26.108	3499.500
Real Estate Development Investment	12	1166.280	1945.230	1581.213	233.821	1626.925
Residential Completed Area	12	1166.280	3472.740	2764.681	355.622	2693.675
Average Sales Price of Commercial Housing	12	4433.000	7023.000	6051.705	1029.855	6743.985
Commercial Housing Sales Area	12	1433.910	3034.920	2064.373	470.619	2160.745

Demand factors: Regional GDP (x_1) reflects economic development and boosts real estate [7]; urban per capita disposable income (x_2) determines purchasing power, with higher income strengthening housing demand; commercial housing sales area (x_3) inversely relates to price; population (x_4) directly affects housing demand.

Supply factors: Residential completed investment (x_5) determines supply quantity, with higher investment increasing supply and lowering prices; residential completed area (x_6) directly reflects commercial housing supply and affects market prices.

In this context, in-depth exploration of the differences in factors affecting housing prices is of great significance for accurately regulating the real estate market and promoting its healthy development. This paper also selects these 7 influencing factors as indicators to analyze the housing prices in Shanxi Province. The data of each indicator are shown in Table 1.

This paper sets the average housing price γ as the dependent variable, and then sets 6 influencing indicators as independent variables for multiple linear regression analysis, establishing the following model:

$$\gamma = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon \quad (1)$$

Where β_1 、 β_2 、 β_3 、 β_4 、 β_5 、 β_6 are regression coefficients, β_0 is a constant term, and ε is the residual, which is independent of the 6 independent variables [2].

To avoid correlation between variables, Qingqi Zhang first used Spearman correlation coefficient to analyze the main factors affecting housing prices, screened out the factors that significantly affect ordinary housing prices and conducted combination analysis [7]. Therefore, here the paper conduct a Pearson correlation analysis on the 6 selected independent variables, as shown in Table 2.

Table 2. Pearson Correlation - Standard Format

	Average sales price of commercial housing (yuan/square meter)
Gross Regional Product(in 100 million yuan)	0.789
Population(in ten thousand people)	-0.834
Commercial Housing Sales Area (in ten thousand square meters)	0.745
Per Capita Disposable Income (for urban residents, in yuan)	0.873
Residential Completed Area (in ten thousand square meters)	0.339
Real Estate Development Investment (in 100 million yuan)	0.636

Table 2 shows average housing price γ has correlation coefficients of 0.789, -0.834, 0.745, 0.873, 0.339, 0.636 with the six factors, with strong correlation except for residential completed area and population [8].

A Multicollinearity test is given. Table 3 shows the first four independent variables have tolerance <0.2 and VIF>5, indicating multicollinearity affecting model estimation. Commercial housing sales area and real estate development investment have tolerance >0.2, with no obvious multicollinearity.

Table 3. Estimated Regression Coefficients Table

.3	Unstandardized Coefficients		Standardized Coefficients	t	p	Collinearity Diagnostics	
	B	Std. Error	Beta			VIF	Tolerance
Constant	-103158.182	179035.709	-	-0.576	0.589	-	-
Gross Regional Product	-0.202	0.15	-1.131	-1.347	0.236	32.678	0.031
Per Capita Disposable Income	0.392	0.236	2.629	1.663	0.157	115.897	0.009
Population	28.478	49.803	0.722	0.572	0.592	73.903	0.014
Residential Completed Area	0.22	1.014	0.076	0.217	0.837	5.68	0.176
Real Estate Development Investment	-0.742	1.255	-0.169	-0.591	0.58	3.764	0.266
Commercial Housing Sales Area	0.51	0.499	0.233	1.022	0.354	2.409	0.415
R squared	0.892						
Adjusted R squared	0.763						
F	F (6,5)=6.893, p=0.026						
Durbin-Watson value	1.387						

A Parameter significance test is given. None of the six independent variables passed the significance test, which was caused by collinearity. Therefore, stepwise regression was used to solve this problem, and Table 4 was created.

Table 4. Stepwise Regression Analysis Table

	Unstandardized Coefficients		Standardized Coefficients	t	p	Collinearity Diagnostics	
	B	Std. Error	Beta			VIF	Tolerance
Constant	1832.337	762.397	-	2.403	0.037*	-	-
Per Capita Disposable Income	0.130	0.023	0.873	5.648	0.000**	1.000	1.000
R squared	0.761						
Adjusted R squared	0.737						
F	F (1,10) = 31.902, p= 0.000						
Durbin-Watson value	0.590						

In the analysis process, five factors with insignificant significance were removed, and a new multiple regression model could be established. The final multiple linear regression equation was written using standardized regression coefficients:

$$\gamma = 0.13x_2 + 1832.337 \quad (2)$$

$$\text{that is } \gamma = 0.13 * \text{ per capita disposable income} + 1832.337 \quad (3)$$

Table 5. The model summary

R	R^2	Adjusted R^2	RMSE	DW	AIC	BIC
0.945	0.892	0.763	323.811	1.387	186.778	190.173

A Variance significance test given As Jae-Jong Kim (2022) did, Table 5 shows the model's R-squared is 0.892, meaning the six factors explain 89.2% of commercial housing price changes.

Table 6. Variance analysis table

	Sum of Squares	df	Mean Square	F	p
Regression	10408366.77	6	1734727.795	6.893	0.026
Residual	1258246.301	5	251649.26		
Total	11666613.07	11			

An Equation significance test given It can be seen from Table 6 that when the model is subjected to F - test, it is found that the model passes the F - test ($F = 6.893$, $p = 0.026 < 0.05$), which means that the model construction is meaningful.

3. Reflection and Prospect on Methods

In this paper, only one data analysis method was used to integrate and analyze the massive and abstract data, so the resulting model may not perfectly conform to real life. To avoid this situation, in future studies, the paper can combine deep learning and use programming languages such as Python to run code for data processing, continuously fit and optimize the model to obtain the best results. Meanwhile, with this study as a foundation, and by continuously revising towards this direction, future research on housing in Shanxi Province will point out the way for people's house-purchasing decisions and what real estate developers should do.

4. Conclusion

Housing prices fluctuate significantly across different regions and time periods due to varying influencing factors. Through regression analysis, it is revealed that residential GDP, completed investment in commercial housing, and completed area of commercial housing are the primary drivers of housing prices in Shanxi Province. With the continuous enhancement of national strength, people's living standards have improved, leading to growing housing demand. This increased demand has spurred real estate developers to expand investment and construction scale. Among these factors, residential commercial housing investment exerts the most profound impact, followed by regional GDP, while the influence of completed area is relatively weaker.

To maintain stable housing prices in Shanxi, the government should implement targeted policies: adhere to the "three stabilities" (stable prices, supply, and expectations) principle, optimize the real estate market structure, accelerate the construction of affordable housing, and advance the long-term mechanism for real estate regulation without relying on it as a short-term economic stimulus. Developers ought to construct housing tailored to different age groups based on residents' needs, and consumers should adopt rational purchasing behaviors to promote healthy market development.

References

- [1] Zhou J, Zhang H, Gu Y J, Pantelous A A. Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai: Published online: 27 February 2018 ©Springer-Verlag GmbH Germany, part of Springer Nature. 2018.
- [2] Dang G Y, Yang T. Multiple linear regression analysis of influencing factors of housing prices in Tangshan City. Journal of Hebei United University (Social Science Edition), 2014, 14 (02): 21-25.

- [3] Wu H, Jiao H Z, Yu Y, Zeng Z, Li Z G, Peng Z H, Liu L B and Zeng Z. Influence Factors and Regression Model of Urban Housing Prices Based on Internet Open Access Data. Received: 24 April 2018; Accepted: 17 May 2018; Published: 22 May 2018.
- [4] Wan L P. Research on influencing factors of second-hand housing prices in Chongqing —— based on the analysis of hierarchical model. *Modern Marketing*, 2021 (04): 112-113.
- [5] Xie R, Du Y, Zhu J M. Econometric analysis of influencing factors on housing prices in Anhui based on multiple regression. *Journal of Natural Science of Harbin Normal University*, 2022, 38 (04): 36-42.
- [6] Zhu Y R, Qian Y H, Guo H P. Statistical inference method for influencing factors of housing prices. *Economic Theory and Practice*, 2022 (05): 58-63.
- [7] Zhang Q Q. Housing Price Prediction Based on Multiple Linear Regression. Received 24 August 2021; Accepted 18 October 2021; Published 29 October 2021.
- [8] Chu Y C, Wang S Y, Chu Y C. Research and prediction on influencing factors of Jinan housing prices —— from the perspective of grey correlation analysis, regression equation and multiple linear regression analysis. *National Circulation Economy*, 2017 (36): 51-54.
- [9] Zhang L H, Analysis and Optimization Research on Influencing Factors of Housing Prices in Dongying City, *Economic Research Guide*, 2023.
- [10] Qin Y J, Comparative Study on Housing Price Prediction Models Based on Multiple Linear Regression and Random Forest Algorithm, *Modern Information Technology*, 2024.