ISSN 2959-6157

Research on Factors Influencing Housing Price based on Principle Component Analysis and Polynomial Regression

Jingcheng Shi[1, *]

[1]College of Science, Shanghai University, Shanghai, 200444, China
*Corresponding author: 1073485684@shu.edu.cn

Abstract:

Housing price is influenced by natural, social, economic and other factors. Only by understanding and conducting specific research on these influencing factors can people grasp the laws of housing price movements. Focusing on this issue, this paper chooses the average unit housing prices and their influencing factors of 11 cities in Zhejiang Province in 2022 as the research objects, then establishes a model reflecting the quantitative relationship between housing price and influencing factors. Firstly, relevant data is collected from official websites, including housing prices and 10 main influencing factor indicators of above cities. Secondly, principal component analysis is used for data preprocessing, and five prominent indicators are selected by calculating the relative contribution rate of each factor. Then, by drawing functional images and solving correlation coefficients, a qualitative analysis is conducted on the basic algebraic structure of the regression model. The results indicate that all independent variables include independent terms and cross-terms. Subsequently, through polynomial regression, data matrices for both dependent variable and independent variables are set to solve the coefficients of constant term, linear terms, quadratic independent terms and quadratic cross-terms respectively. Therefore, a model of factors influencing housing price is established. Finally, residuals sum of squares, mean square error, regression sum of squares and determination coefficient are calculated in sequence for the regression model. It turns out that all the statistics are within the ideal range, indicating high precision of data fitting and verifying the accuracy and reliability of the model.

Keywords: Housing price; influencing factor; principal component analysis; polynomial regression.

1. Introduction

Housing price refers to the market value of a building together with the land it occupies for a specific time. In the whole price system, it always occupies a fundamental and leading position, which is of great significance to both national economic and social development. In reality, housing price level is influenced by a series of factors, including natural factors, social factors, economic factors, etc. In order to deeply understand the impact of these factors and conduct quantitative research on it, this paper chooses the average unit housing prices and their influencing factors of 11 cities in Zhejiang Province in 2022 as the research objects. By selecting the main factors, this paper will establish a model reflecting the quantitative relationship between housing prices and these factors, so as to grasp the law of housing price changes.

In fact, many experts and scholars have explored the factors influencing housing prices in specific regions before. Using various mathematical modeling and statistical analysis methods, they established corresponding models of housing price influencing factors. In 2016, Zhou et al. collected the housing price data of Jiangxi Province in the past 9 years and established the housing price regression model by using the step-by-step linear regression method according to four main influencing factors [1]. Meanwhile, based on the average housing prices of 15 cities, Chen et al. used grey relational degree model to calculate the correlations between several influencing factors and housing prices, thus establishing the GM (1,1) model for housing price prediction [2]. In 2017, Wang et al. focused on the change of housing price in Kunming and established an influencing factors model of housing price based on the multiple linear regression method, which was improved by the quantitative correlation analysis [3]. Meanwhile, Chen et al. conducted a study on the housing price fluctuation in Xiamen by using the grey relational analysis method [4]. In 2018, Li et al. established a multiple linear regression model for the housing price of Sanya. They introduced dummy variables and adopted time series analysis to study the housing price fluctuations before and after the introduction of housing purchase restriction policy [5].

Meanwhile, Qi et al. built a Malthus model of housing price for floating population based on GM (1,1) grey system model and time series analysis. Then they introduced some parameters to establish a BP neural network model of influencing factors of housing price [6]. In 2019, Bao G. et al. created an evaluation system of the impact of real estate regulation policies on housing prices in Hainan Province, which was based on improved particle swarm optimization algorithm and fuzzy analytic hierarchy process. Furthermore, they used multiple linear regression analysis method to predict the commercial housing prices on a monthly basis [7]. Meanwhile, Luo et al. selected the sample data of Fuzhou in recent 8 years and built a quantile regression model of housing price based on dynamic panel data by studying the specific impact mechanism of factors on local housing price [8]. In 2020, Dai et al. analyzed main factors considered by second-hand house buyers and combined them with the second-hand house data of Chengdu, thus establishing a multiple linear regression model which affected the housing price per unit area [9]. Meanwhile, based on the housing price data of Shenzhen in the past 20 years, Li et al. chose ten indicators affecting the local housing price from three dimensions. They used grey correlation degree model to explore the correlation degree between the prices of commercial housing in Shenzhen under one-child policy and two-child policy [10].

Undoubtedly, these results have important reference value for the judgment of the factors influencing housing price. However, the author finds that many research methods and processes still have some deficiencies. Firstly, the research objects are often too large in scale, thus neglecting the horizontal differences within the region. Therefore, the division of research objects may be too rough, leading to large deviations in research results. Secondly, affected by special public events such as the corona-virus epidemic, some annual housing price data fluctuated greatly, which deviated from the normal operating trend. Therefore, if the data set is selected based on temporal order rather than spatial difference, the fitting accuracy of the established model may decrease while the data error will increase. Thirdly, most studies lack the step of data preprocessing. In fact, although there exist factors that affect housing price, their impact is not significant. Taking them into consideration will increase the subsequent modeling workload and algorithm complexity, making it difficult to highlight the focus and weakening the contribution of core indicators. Fourthly, the data fitting methods used in many studies are traditional multiple linear regression methods. However, before in-depth analysis, it is impossible to determine that there exists a linear relationship between housing price and any of its influencing factors, not to mention that there is no direct correlation between any two influencing factors. If linear regression is rashly applied here, it may lead to an inaccurate reflection of the relationships between these variables. Meanwhile, since the highest power of the corresponding expression in the linear regression method is only 1, the fitting accuracy will reduce and the data error rate will increase, thus weakening the reference significance and application value of the model. Fifthly, many studies lack the basic step of model testing, which will lead to a lack of foundation for model evaluation and application. Although some studies use grey relational analysis and Markov chain analysis, they neglect the fact that these two methods are only applicable within a certain range. Actually, neither of them is suitable for testing housing price models.

2. Methods

2.1 General Description

This paper focuses on the quantitative relationship between the unit housing price level and its influencing factors in Zhejiang Province in 2022. Considering the limitations of previous similar studies, this paper collects values of relevant indicators from official websites, and uses principal component analysis as the data screening method. Based on a reasonable judgment of the linear relationship between each indicator variable, a polynomial regression method is used as the data fitting method to establish the model of housing price factors of Zhejiang Province. Finally, various statistical parameters are used to conduct quantitative data testing on the obtained model. The whole process is supported by Python programming.

2.2 Indicator Selection

Through comprehensive analysis, the paper selects a series of factors affecting the average unit housing price from the aspects of nature, society and economy, and then analyzes their mechanism and effect.

The paper selects a series of factors affecting the average unit housing price in of Zhejiang Province from the aspects of nature, society, and economy through comprehensive consideration, and briefly analyzes their mechanisms and effects.

2.2.1 Natural Factors

Natural factors have many influences on housing price, which can directly or indirectly affect the current price and long-term value of real estate. The indicators of natural factors selected in this study include urban land area and average altitude, which are denoted as x_1 and x_2 respectively.

2.2.2 Social Factors

The influences of social factors on housing price are

complex and diversified, including supply and demand relationship, policy adjustment, industrial structure and so on. The interaction of these factors together affects the basic trend of housing price and the overall market performance. The social factor indicators selected in this study include permanent resident population, natural population growth rate and urbanization rate, which are denoted as x_3, x_4 and x_5 respectively.

2.2.3 Economic Factors

Economic factors are the most important and core factors in terms of determining the level of housing price. They directly affect the supply and demand balance of the real estate market, especially the unilateral demand situation. The indicators of economic factors selected in this study include the gross regional product, the balance of residents' deposits at the end of the year, the amount of investment in real estate development, the completed area of commercial housing and the sales area of commercial housing, which are recorded as x_6, x_7, x_8, x_9 and x_{10} respectively.

Through the above process, table 1 including indicators of housing price and its influencing factors is established, which is shown as follows:

Table 1. Housing price and its influencing factors of Zhejiang Province in 2022

Indicator variable	Practical meaning	Indicator property
x_1	urban land area (square kilometer)	natural factor
x_2	average altitude (meter)	natural factor
x_3	permanent resident population (thousand people)	social factor
x_4	natural population growth rate (%)	social factor
x_5	urbanization rate (%)	social factor
x_6	gross regional product (billion RMB)	economic factor
x_7	balance of residents' deposits at the end of year (billion RMB)	economic factor
x_8	amount of investment in real estate development (billion RMB)	economic factor
x_9	completed area of commercial housing (thousand square meter)	economic factor
x_{10}	sales area of commercial housing (thousand square meter)	economic factor
y	average housing price per square meter (RMB per square meter)	-

2.3 Data Acquisition

Through consulting authoritative official websites such as Zhejiang Provincial Statistical Yearbook and Zhejiang Provincial Statistical Bulletin on National Economic and Social Development, the author obtains the average unit housing price data of 11 prefecture-level cities in Zhejiang Province in 2022 as well as the data of the aforementioned 10 factors affecting housing price.

2.4 Method Introduction

In the following part, the paper will focus on screening of indicator variables, analysis of data properties, establishment of influencing factors model and test of this model. To accomplish these steps, two key methods, principal component analysis and polynomial regression, will be used.

2.4.1 Principal Component Analysis

Principal component analysis is an important dimensionality reduction method. In reality, people want to obtain as much information as possible by using as few variables as possible. In fact, there is a certain correlation among many original variables, and the information reflected by them also has a certain overlap. The core task of principal component analysis is to delete the repetitive content of the highly correlated variables according to the specific information contribution rate so as to establish as few new variables as possible. These new variables are the linear combination of original variables, which are not only independent of each other, but also can accurately reflect most of the information contained in original variables.

2.4.2 Polynomial Regression

Regression analysis is a type of statistical analysis method that uses the principles of mathematical statistics to conduct quantitative research on a large number of data and determine the correlation between dependent variables and specific independent variables. A regression equation with good correlation can not only reflect the specific relationship between the dependent variable and its independent variables, but also extrapolate it to predict the future trend of the dependent variable. Therefore, it has import-

ant theoretical and practical significance.

3. Result and Discussion

3.1 Screening Of Influencing Factors

3.1.1 General Description

To simplify the process of establishing the model of factors influencing housing price and make it more concise and targeted, this chapter considers to adopt principal component analysis. With the help of Python programming, this paper quantitatively analyzes the impact of various original indicators on housing price and calculates their relative contribution rates respectively. Based on these rates, the indicators ranked in the top half are selected and retained in order. Relatively speaking, these indicators have a more significant impact on housing price.

3.1.2 Results Realization

Based on specific steps of principal component analysis, this paper obtains the descending ranking results of relative contribution rates of all original indicator variables, which are shown in Table 2:

Table 2. The relative contribution rates of all factors influencing housing price

Influencing factor	indicator variable	Relative contribution
gross regional product	x_6	11.4791%
balance of residents' deposits at the end of year	x_7	11.3963%
amount of investment in real estate development	x_8	11.1806%
permanent resident population	x_3	11.1093%
urban land area	x_1	11.0437%
completed area of commercial housing	x_9	11.0174%
sales area of commercial housing	x_{10}	10.4905%
natural population growth rate	x_4	9.5470%
urbanization rate	x_5	9.1836%
average altitude	x_2	3.5524%

From the table above, it is obvious that among all factors affecting housing price, the independent relative contribution rates of the top 50% are gross regional product, balance of residents' deposits at the end of year, amount of investment in real estate development, permanent resident population and urban land area respectively, and the corresponding indicator variables are x_6, x_7, x_8, x_3, x_1. By retaining the indicator variables which are most important, it will not only make the key factors more prominent, but also simplify the subsequent modeling process. The influencing factors of housing price after quantitative screening are shown in the following table 3:

Table 3. The factors of housing price after being screened by principal component analysis

Influencing factor	indicator variable
gross regional product	x_6
balance of residents' deposits at the end of year	x_7
amount of investment in real estate development	x_8
permanent resident population	x_3
urban land area	x_1

3.2 Model Structure Analysis

3.2.1 General Description

This chapter considers to achieve qualitative analysis of the basic algebraic structure of the polynomial regression model to be built. Since the correlation coefficient of each variable in the original problem can only reflect the algebraic relationship between two indicators, this paper

chooses to set the highest power of the polynomial as 2. Therefore, the regression model fitted is at most a quadratic polynomial.

In terms of specific algebraic structure, to begin with, it is necessary to qualitatively judge whether there are independent variables in the regression model while each of them contains only linear term, i.e. the highest power of the independent variable is 1. It can be realized by drawing function images between the dependent variable and their independent variables in turn. If it is found that the image of a function is basically a straight line rather than an ordinary curve, it means that there exists a linear relationship between the dependent variable and the independent variable. Hence, the corresponding independent variable only involves linear term, otherwise it involves quadratic term. In addition, it is necessary to qualitatively judge whether there are independent variables without cross-terms in the regression model, i.e. there exists no correlation between any two independent variables. It can be achieved by calculating the correlation coefficient among all independent variables. If the correlation coefficient between two independent variables is close to zero, it means that there is almost no correlation between the two, and no quadratic cross-term is involved. Otherwise, it will involve the quadratic cross-term. What's more, since the regression model must include all independent variables, their power must be 1 at least. Meanwhile, since the correlation coefficient of any independent variable and itself is 1, all independent variables should involve quadratic independent terms. Through above steps, this paper can determine the number of linear terms of each independent variable, the number of quadratic cross items it involves, and the relationship with any other independent variable. By adding these quadratic independent terms, the basic algebraic structure of the polynomial can be understood.

3.2.2 Number of Linear Terms

Through principal component analysis, this paper has obtained the outstanding influencing factors of housing price, including gross regional product, balance of residents' deposits at the end of year and amount of investment in real estate development, permanent resident population and urban land area. Based on relevant data of 11 cities in Zhejiang Province, the average housing price is taken as the dependent variable, and the factors affecting housing price are taken as the independent variables. Then functional images are drawn respectively according to the relative contribution rate of each factor. If the functional image of the dependent variable about an independent variable is almost a straight line, it can be considered that the relationship between the two is basically linear, so that the independent variable only has a linear term without any quadratic term. On the contrary, it is considered that the relationship between the two is non-linear, and the independent variable necessarily involves quadratic terms and may also involve a linear term at the same time. The images of each function relationships are listed as follows (Figure 1, 2, 3, 4 and 5):

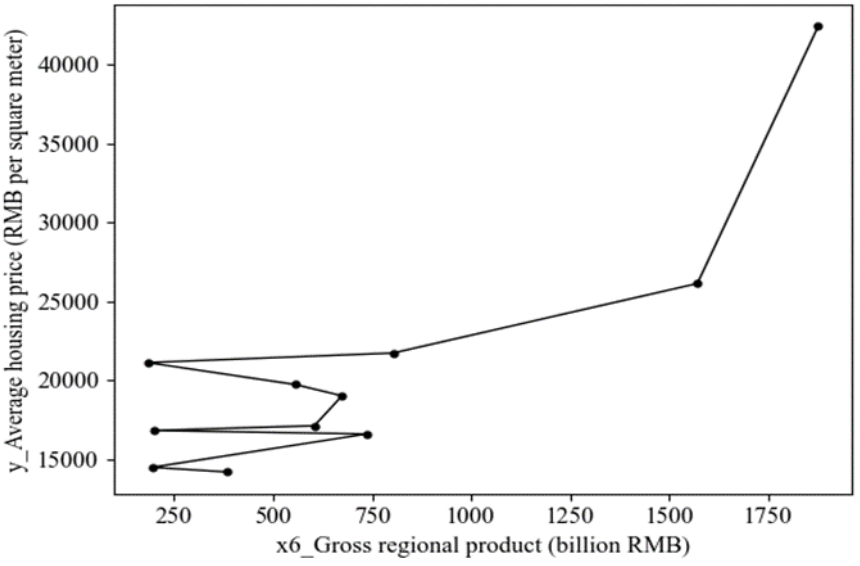

Fig. 1 Functional relationship between average housing price and gross regional product of prefecture-level cities in Zhejiang Province in 2022

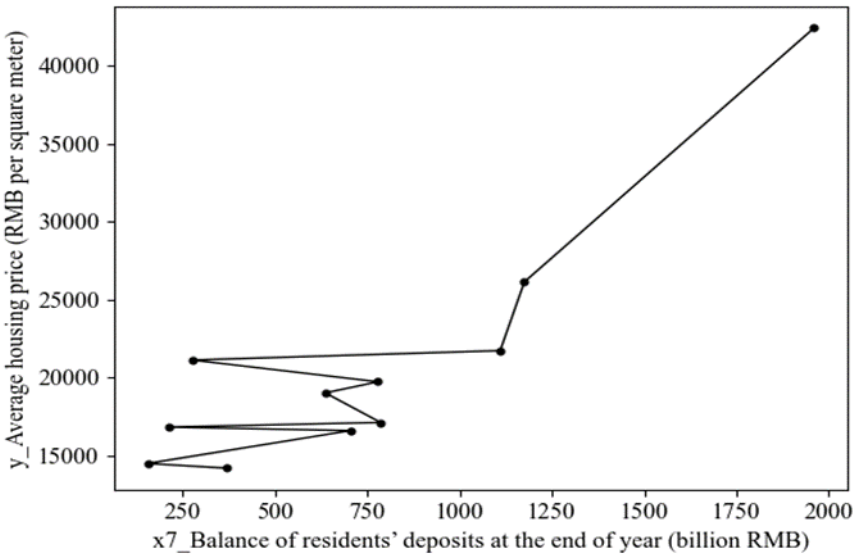

Fig. 2 Functional relationship between average housing price and balance of residents' deposits at the end of year of prefecture-level cities in Zhejiang Province in 2022

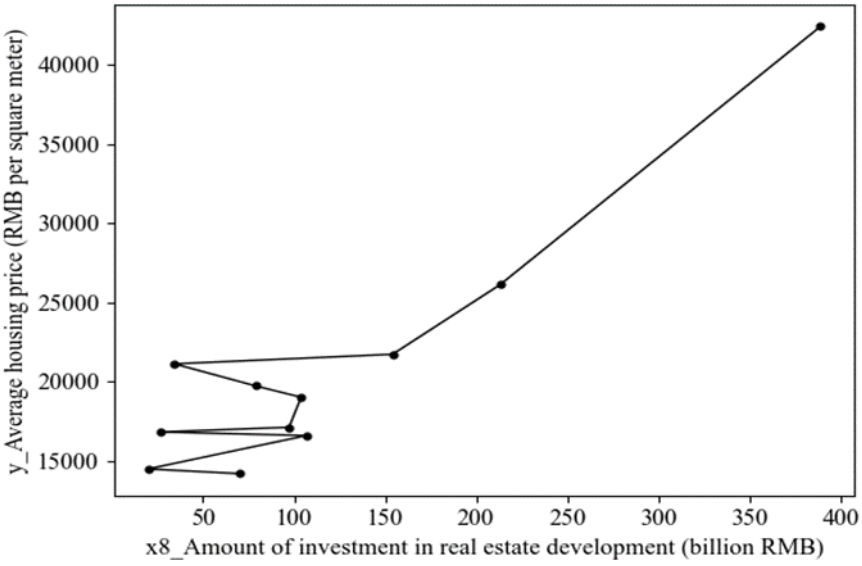

Fig. 3 Functional relationship between average housing price and amount of investment in real estate development in prefecture-level cities of Zhejiang Province in 2022

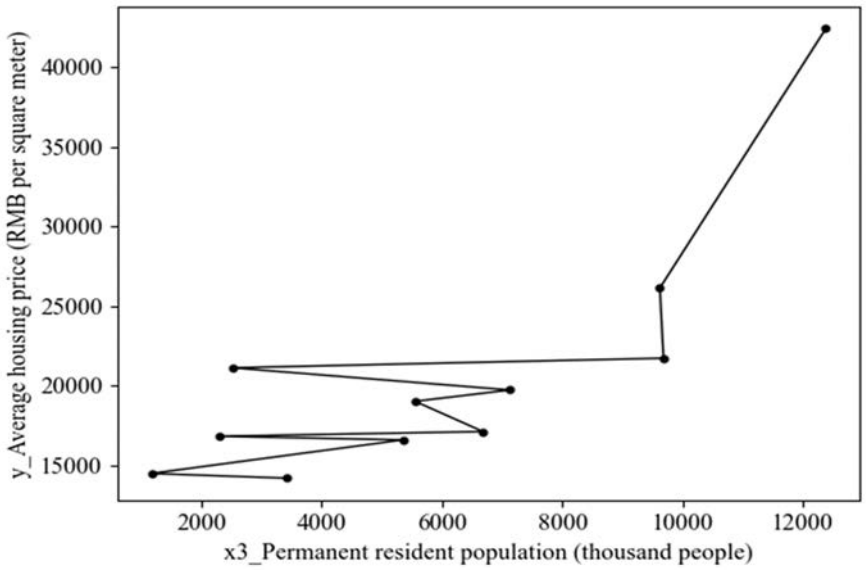

Fig. 4 Functional relationship between average housing price and permanent resident population of prefecture-level cities in Zhejiang Province in 2022

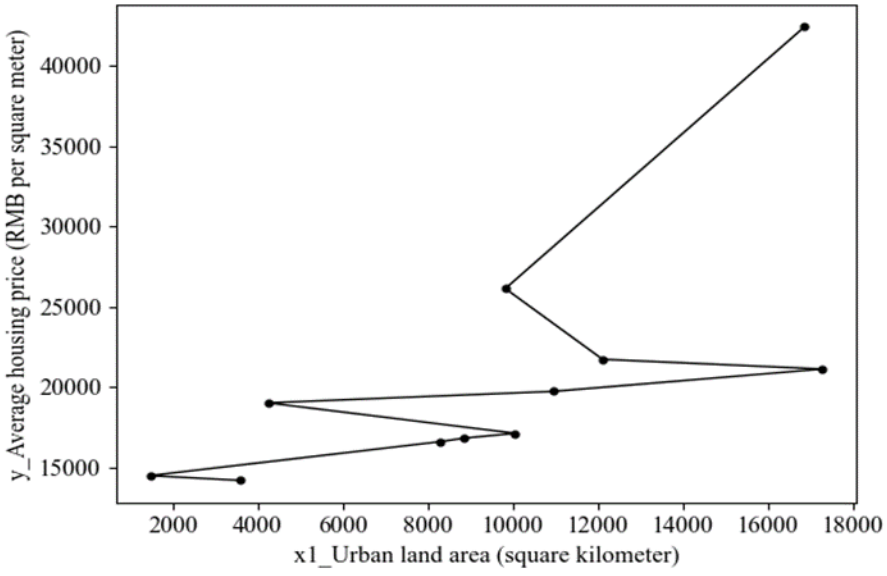

Fig. 5 Functional relationship between average housing price and urban land area of prefecture-level cities in Zhejiang Province in 2022

It can be seen that the functional images of the dependent variable about their variables are ordinary polylines instead of straight lines, so the relationship between the dependent variable and any independent variable is non-linear. Meanwhile, all the independent variables in the polynomial regression model not only involve linear terms, but also involve the quadratic term, i.e. they all contain the basic algebraic structure,

$$ax_i^2 + bx_i + c\left(1 \leqslant i \leqslant 10, a, b, c \in \mathbb{R}, a \neq 0\right). \quad (1)$$

3.2.3 Number of quadratic cross-terms

Based on the relevant data from the aforementioned 11 cities, this paper will calculate the correlation coefficient between any two of the five independent variables after being screened. If the correlation coefficient is close to zero, it can be considered that there exists basically no correlation between the two, and the polynomial regression model does not contain the quadratic cross-term of them. Otherwise, the quadratic cross-term is involved. The list of specific correlation coefficients is shown as follows (Table 4):

Table 4. Correlation coefficients of the factors influencing housing price

indicator variable	x_6	x_7	x_8	x_3	x_1
x_6	1.0000	0.9428	0.9627	0.9039	0.4052
x_7	0.9428	1.0000	0.9702	0.9710	0.5402
x_8	0.9627	0.9702	1.0000	0.8982	0.4886
x_3	0.9039	0.9710	0.8982	1.0000	0.5174
x_1	0.4052	0.5402	0.4886	0.5174	1.0000

It can be seen from the above table that among all five independent variables screened by principal component analysis, the correlation coefficients between any two are all strictly greater than zero, which indicates that there exists a certain association relationship between any two independent variables. Therefore, all the independent variables in the polynomial regression model involve quadratic cross-terms, i.e., they all contain the basic algebraic structure,

$$dx_i x_j \left(1 \leqslant i, j \leqslant 10, i \neq j, d \in \mathbb{R}, d \neq 0\right). \tag{2}$$

3.3 Model Establishment

3.3.1 General description

The basic algebraic structure of the established polynomial regression model has been analyzed qualitatively, including the number of constant terms, linear terms, quadratic independent terms and quadratic cross-terms of each variable and their relations. The following part will focus on the screened independent variables and solve the specific expression of the model from the quantitative point of view, involving the set and solution of each coefficient. Based on this, a complete model of factors influencing average unit housing price in Zhejiang Province in 2022 will be established.

3.3.2 Model preparation

Through the algebraic structure analysis of the polynomial regression model, it can be observed that the indicator variables obtained after being screened by principal component analysis all include a linear term and a quadratic independent term. Moreover, there exist quadratic cross-terms between any two indicator variables. Meanwhile, there is inevitably a constant term in the model. Since the highest power of the model is 2, then it belongs to a class of complete quadratic polynomial models, which can be set in the following algebraic form:

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_{11} x_1^2 + \\ & \beta_{13} x_1 x_3 + \beta_{16} x_1 x_6 + \beta_{17} x_1 x_7 + \beta_{18} x_1 x_8 + \beta_{33} x_3^2 + \\ & \beta_{36} x_3 x_6 + \beta_{37} x_3 x_7 + \beta_{38} x_3 x_8 + \beta_{66} x_6^2 + \beta_{67} x_6 x_7 + \\ & \beta_{68} x_6 x_8 + \beta_{77} x_7^2 + \beta_{78} x_7 x_8 + \beta_{88} x_8^2 \end{aligned} \tag{3}$$

where the dependent variable is y, and the constant term is β_0. There are 5 linear terms, which are set as x_1, x_3, x_6, x_7, x_8 respectively, and their coefficients are successively set as $\beta_1, \beta_3, \beta_6, \beta_7, \beta_8$. There are 5 quadratic independent items, which are set as $x_1^2, x_3^2, x_6^2, x_7^2, x_8^2$ respectively, and their coefficients are successively set as $\beta_{11}, \beta_{33}, \beta_{66}, \beta_{77}, \beta_{88}$. There are 10 quadratic cross items, which are set as $x_1x_3, x_1x_6, x_1x_7, x_1x_8, x_3x_6, x_3x_7, x_3x_8, x_6x_7, x_6x_8, x_7x_8$ respectively, and their coefficients are successively set as $\beta_{13}, \beta_{16}, \beta_{17}, \beta_{18}, \beta_{36}, \beta_{37}, \beta_{38}, \beta_{67}, \beta_{68}, \beta_{78}$. Hence the polynomial has 21 terms altogether, whose coefficients will be solved individually.

3.3.3 Model solution

To begin with, the original data is divided into a dependent variable dataset and an independent variable dataset. The dependent variable dataset includes the average housing price of 11 cities in Zhejiang Province in 2022, with the unit being RMB per square meter. While the independent variable dataset includes the gross regional product, balance of residents' deposits at the end of year, amount of investment in real estate development, permanent resident population and urban land area for these cities in the same year, with the units being billion RMB, billion RMB, billion RMB, thousand people and square kilometer respectively. Therefore, the dependent variable dataset contains 11 rows and 1 column, which can be represented the column vector $Y \in \mathcal{F}^{11}$. The independent variable dataset contains 55 data in 11 rows and 5 columns, which can be expressed as the matrix $X \in \mathbb{R}^{11\times5}$. Then substitute the original data, it can be derived that Y=(42445, 26136,

21728, 21123, 19728, 19015, 17117, 16818, 16593, 14482, 14184)^T .

$$X = \begin{pmatrix} 1875.30 & 1959.88 & 388.91 & 12376.00 & 16850.00 \\ 1570.40 & 1174.98 & 213.17 & 9618.00 & 9816.00 \\ 803.00 & 1109.59 & 154.07 & 9679.00 & 12103.00 \\ 183.10 & 276.02 & 34.33 & 2515.00 & 17275.00 \\ 556.20 & 778.53 & 78.84 & 7127.00 & 10942.00 \\ 673.90 & 635.92 & 103.87 & 5551.00 & 4237.00 \\ 604.10 & 785.54 & 97.24 & 6678.00 & 10050.00 \\ 200.30 & 211.87 & 26.56 & 2290.00 & 8845.00 \\ 735.10 & 704.12 & 106.79 & 5353.00 & 8279.00 \\ 195.10 & 154.41 & 19.79 & 1170.00 & 1459.00 \\ 385.00 & 367.87 & 70.41 & 3413.10 & 3580.00 \end{pmatrix}. \quad (4)$$

By using data fitting method, the coefficient values of each item in the polynomial regression model can be obtained. The specific results are listed as follows (Table 5):

Table 5. Coefficient values of the polynomial regression model

Coefficient name	Coefficient attribute	coefficients value
β_0	constant term	1.4102×10^{4}
β_1	linear item	9.3568×10^{-5}
β_3	linear item	8.9841×10^{-5}
β_6	linear item	2.0203×10^{-5}
β_7	linear item	1.2484×10^{-3}
β_8	linear item	6.3016×10^{-4}
β_{11}	quadratic independent item	7.5386×10^{-3}
β_{33}	quadratic independent item	3.1853×10^{-2}
β_{66}	quadratic independent item	5.1172×10^{-3}
β_{77}	quadratic independent item	4.8419×10^{-5}
β_{88}	quadratic independent item	7.4262×10^{-5}
β_{13}	quadratic cross-term	3.8064×10^{-2}
β_{16}	quadratic cross-term	1.5096×10^{-2}
β_{17}	quadratic cross-term	-1.0647×10^{-2}
β_{18}	quadratic cross-term	3.8069×10^{-3}
β_{36}	quadratic cross-term	1.2702×10^{-2}
β_{37}	quadratic cross-term	3.2388×10^{-3}
β_{38}	quadratic cross-term	-6.6471×10^{-3}
β_{67}	quadratic cross-term	9.0913×10^{-3}
β_{68}	quadratic cross-term	-2.4913×10^{-2}
β_{78}	quadratic cross-term	3.6851×10^{-4}

As can be seen from the above table, the specific form of the model for the influencing factors of the average unit

housing price level of cities in Zhejiang Province in 2022 is shown as follows:

$$y = 1.4102 \times 10^4 + 9.3568 \times 10^{-5} x_1 + 8.9841 \times 10^{-5} x_3 + \ldots + 7.4262 \times 10^{-5} x_8^2 \quad (5)$$

3.4 Model Test

3.4.1 General description

Based on the polynomial regression method, the coefficients of each constant term, linear term, quadratic independent term and quadratic cross-term are accurately solved by setting independent variable matrix and dependent variable matrix respectively. This leads to the establishment of the influencing factors model for the average unit housing price of 11 cities in Zhejiang Province in 2022. The following will introduce some important statistics, including residual sum of squares, mean square error, regression sum of squares and determination coefficient. These corresponding statistics are calculated successively for the resulting polynomial regression model. This approach not only achieves a quantitative evaluation of the rationality and accuracy of the model, but also effectively analyzes the specific errors in data fitting, thereby providing a more comprehensive understanding of the model.

3.4.2 Result realization

According to the polynomial regression model established above, the corresponding four statistics are calculated in order to examine the fitting accuracy of the model coefficients and the magnitude of errors of the whole model. The results are shown as follows:

Table 6. The statistics indicators of the polynomial regression model

Statistics indicator	Specific meaning	Numerical value
SSres	sum of squares of residuals	1.6018×10^{-20}
MSE	mean square error	1.4561×10^{-21}
SSreg	regression of Sum of Squares	6.3306×10^{8}
R^2	coefficient of determination	1.0000

From the above table 6, it can be seen that the values of residual sum of squares and the mean square error of the polynomial regression model are both extremely small, which are almost approaching zero. The value of the regression sum of squares is quite large, with an order of magnitude reaches 10^8. The value of the coefficient of determination is almost equal to 1. Hence it can be concluded that the regression model has high precision and low data fitting error, which verifies the accuracy and reliability of it to some extent.

4. Conclusion

This paper focuses on the issue of housing price, and selects the average unit housing prices and their influencing factors of 11 cities in Zhejiang Province in 2022 as the research objects. By establishing a model, the quantitative relationship between housing price and its influencing factors has been revealed.

In the process of consulting relevant studies, it has been found that similar studies generally have the common problems, including rough division of research objects, neglecting the influences of special years. Meanwhile, the problems of lacking data preprocessing and model testing steps, and hasty use of linear regression method without empirical analysis also exist. Therefore, this study tries to make some improvements for the above aspects.

In terms of research steps, firstly, this paper collects data from official websites such as Zhejiang Provincial Statistical Yearbook and Zhejiang Provincial Statistical Bulletin on National Economic and Social Development. The relevant data includes the average housing prices of 11 prefecture-level cities in Zhejiang Province in 2022, as well as 10 influencing factors indicators in natural, social and economic aspects. Secondly, the dimension reduction method, i.e. principal component analysis is used for data preprocessing. By quantitatively calculating the relative contribution rate of each factor, five main indicators are selected, including the gross regional product, resident deposit balance at the end of the year, real estate development investment amount, permanent resident population and urban land area. Thirdly, according to the definition of correlation coefficient, the highest power of the regression model is set as 2. Meanwhile, through functional image drawing and correlation coefficient solving, qualitative analysis is conducted on the basic algebraic structure of the regression model to be established. The results indicate that all selected independent variables involve both linear terms and cross-terms. Hence the quantities and interrelationships of constant terms, linear terms, quadratic independent terms and quadratic cross-terms in the model are preliminarily determined. Fourthly, by employing the data fitting method of polynomial regression, the coefficients

of the aforementioned terms are set and solved respectively, which is based on setting both dependent variable data matrix and independent variable data matrix. Therefore, a model reflecting the quantitative relationship between the average unit housing price of Zhejiang Province in 2022 and its five influencing factors is established. Finally, key statistics such as residual sum of squares, mean square error, regression sum of squares and determination coefficient are introduced. They are calculated successively for the established polynomial regression model. The result shows that all values are within the ideal range, indicating high model accuracy and small data fitting error of this model, thereby verifying its reliability and effectiveness to some extent.

References

[1] Zhou Ermin, Zhu Jin, Wang Guiyong, et al. Construction and empirical analysis of housing price influencing factors model: A case study of Jiangxi Province. Journal of Lanzhou University of Finance and Economics, 2016, 32(04): 34-43.

[2] Chen Qingxin, Wen Qinglan, Lu Haihua, et al. Based on the price influencing factors of grey correlation degree analysis. Modern Marketing (Management), 2019, (05): 66-68.

[3] Wang Pengfei, Yu Kaichao. Analysis of influencing factors of housing price in Kunming City based on multiple linear regression. Software, 2018, 39(09): 152-157.

[4] Chen Ying, Ge Yingqi. Grey correlation analysis of influencing factors of housing price fluctuation in Xiamen: Based on data from 2000 to 2016. Wuyi College Journal, 2018, 05(12): 22-27.

[5] Li Dan, Zhu Jiaming, Li Wei, et al. Study of prices influencing factors based on multivariate regression model. Journal of Liaoning University of Technology (Natural Science Edition), 2019, 33(03): 206-210.

[6] Qi Wenbin, Hou Zongrun, Li Guixi, et al. Analysis and prediction of four-city housing prices based on GM(1,1) model and BP neural network. Computer Knowledge and Technology, 2018, 14(13): 179-185.

[7] Bao Gang, Wu Tingting, Fang Yuan, et al. Study on the impact of purchase restriction policy on housing price based on improved PSR model: A case study of Hainan Province. Mathematics in Practice and Understanding, 2019, 49(20): 296-302.

[8] Huang Jinjin, Luo Jinyan, Liu Jia, et al. Analysis of influencing factors of housing price in Fuzhou City based on quantile regression model of dynamic panel data. Straits Science, 2019, (05): 61-65.

[9] Dai Lei, Li Xueting. Analysis of influencing factors of second-hand housing price based on multiple linear regression model: A case study of a district in Chengdu. Henan Building, 2019, (05): 80-82.

[10] Li Binsheng, Liang Yuhan. An analysis of influencing factors of housing price in Shenzhen under the two-child policy: Based on grey correlation degree. Journal of Shaoguan University, 2020, 41(04): 59-64.