

Regional Economic Prediction Model and Empirical Research Based on Big Data Technology

Yulin Luo^{1,a}, Tianming Huang^{*2,b}

¹Business School Wuchang University of Technology Wuhan, China

²College of Arts and Law Wuhan Donghu University Wuhan, China

Abstract—in order to test the prediction effect of big data technology on regional economy, this paper uses Engle-Granger test to explore the relationship between house prices in Xiangyang City and big data represented by search index. This paper selects the house price data from April 2015 to February 2020, and uses the Baidu Index with a correlation coefficient greater than 0.5 to build a search index, so as to test it empirically. The empirical results show that big data has a strong prediction for house prices in small and medium-sized cities, and the search index constructed by the weight of correlation coefficient is better.

1 Introduction

As we all know, house price is an important means of government macroeconomic regulation and control. The stability of house price affects the stable growth of economy to a certain extent. Therefore, there are many domestic research literatures on house price [1]. Liu Liya et al. (2005) used the co integration method to measure the reasonable level of house price in Shanghai, and analyzed the degree and reason of the maladjustment of house price in Shanghai. Zhang Yali (2011) used the real estate price determination model to draw the conclusion that the real per capita disposable income, expected income and expected real estate rate of return are the main factors for the continuous rise of house price Wang Wenwen et al. (2014) studied the influence of population structure, income distribution, economic policy and other variables on house prices with provincial panel data. To sum up, most of the research literature on housing prices is concentrated in large and medium-sized cities, while little attention is paid to small and medium-sized cities. However, in recent years, the house price in Xiangyang has soared from 5125 Yuan in April 2015 to 9177 Yuan in February 2020, with an increase of 80%, ranking first in the country. So, how to analyze and predict the changes of house prices in small and medium-sized cities?

In 2009, Nature magazine published an article using Google search log to predict the trend of influenza in real time, which quickly attracted the attention of the society to big data. In recent years, people continue to tap the huge potential behind big data, apply it to various fields such as medical treatment, transportation, commercial promotion, and even call it "digital gold" vividly [2]. In academic circles, the ability of big data to predict the economy is gradually put in a prominent position. Zhang Yihao et al. (2014) found that Internet search can be used to predict the

stock market, so as to construct a portfolio to obtain excess returns; Liu taoxiong et al. (2015) discussed whether Internet search can be used to predict the macro-economy by comparing the structured data from the government and the unstructured data from Internet search; Wang Na (2016) proved it by using Baidu search index and media index Big data can improve the accuracy of carbon emission prediction.

Because big data has better prediction ability in the economic field, this paper attempts to explore the relationship between big data and real estate, hoping to provide a way for the economic prediction of small and medium-sized cities.

2 Theoretical analysis of the influence of search index on house price

From an economic point of view, housing is also a kind of commodity, so its price must be determined by its supply and demand. From the perspective of supply, the supply of houses depends on the supply of land to a large extent, and land, as a special commodity, is inelastic [3]. Therefore, in the short term, housing prices are mainly affected by housing demand. In terms of housing demand, consumer demand mainly includes investment demand and speculative demand. Among them, investment demand reflects the rigid demand of consumers and is affected by economic fundamentals and other factors. The demand for speculation depends on the expected return of the house. If the expected return on house price is greater than the general return on social capital, the demand for speculation will increase; otherwise, the demand for speculation will decrease.

Network search is a means of information mining, which can not only obtain information but also record the dynamic of investors. Investors often do not search for information that has nothing to do with it or is not

^aLyl_whxy@163.com

^{b*} Corresponding author: 1660104131@qq.com

interested in it. For example, when an investor searches for mortgage rates, it often indicates that the investor is paying attention to the real estate market, or that it has strong willingness to buy houses. Based on this, this paper believes that fully mining the information of network search can understand the needs of investors, so as to predict the changes of house prices in the future. However, there is no generally accepted view about the influence mechanism of the two. Combining the theory of asset price determination and the theory of behavioral economics, this paper puts forward the following theoretical analysis.

Internet search is actually a reaction of investor sentiment and investor concern. When the investor's sentiment is high, the attention to the asset will inevitably rise, thus showing a higher amount of Internet search; when the investor's sentiment is low, it means that the investor's interest in the asset is weakened, and the investor's attention will be reduced, thus showing a smaller amount of Internet search. According to the relevant theory of behavioral economics, investor sentiment and investor concern are the main factors that affect investor preference and decision-making, and preference and decision-making will cause demand changes to a certain extent, resulting in price fluctuations. Specifically, there are investors and speculators in the real estate market. When the investor's sentiment is high and the attention is rising, the buyer often makes a judgment that the real estate market will have the excess yield in the future, thus affecting the investor's preference and decision-making[4].

At this time, the investors will show a stronger willingness to buy, a shorter time to wait for trading, and more cost consideration they are willing to pay, which will lead to an increase in investment demand; while the speculators will show an increase in the possibility of entering the housing market, the number of transactions, and the majority of buyers, which will lead to an increase in speculative demand. As a result, overall demand in the housing market has risen. On the contrary, when investors are depressed and concerned about the decline, it often means that the probability of the expected future real estate market yield decline is enhanced, which affects investors' preference and decision-making. At this time, the investor's willingness to buy will be reduced, the time to wait for trading will be prolonged, and the cost consideration willing to pay will be reduced, which will lead to the reduction of investment demand; while the possibility of speculators entering the market will be reduced, the number of transactions will be reduced, and the trading direction is mostly selling, which will lead to the reduction of investment speculation demand. As a result, the overall demand for housing has declined. According to the theory of supply and demand, the supply of short-term real estate market is not easy to change. The increase of housing demand will lead to the rise of house prices, while the decrease of demand will lead to the fall

of house prices. To sum up, the use of Internet search can predict the future changes in housing demand, so as to predict the future changes in house prices.

3 Empirical test

3.1 Data source and variable selection

1. House price index P_1 . The original house price data of this paper comes from Xiangyang real estate information network, which collects the monthly average of the transaction price of residential commercial housing in Xiangyang city. Comparing the data with the monthly benchmark data of Xiangyang house in wind database, we find that the correlation between the two is as high as 0.98, so we think the data is accurate and reliable. This paper selects 59 monthly data from April 2015 to February 2020, and sets the house price in April 2015 to 100, and processes the data year-on-year to get the house price index.

2. The construction of search index. Based on the experience of references at home and abroad, this paper uses Baidu Index to represent the Internet search data of investors. Limited the search area to Xiangyang area, and used Python software to crawl the Baidu Index of 14 key words related to house price from April 2015 to February 2020. The selected keywords are commonly used words directly or indirectly related to house price, including house price, house purchase, precautions for house purchase, provident fund, provident fund query, deposit interest rate, home decoration, property tax, house loan, house loan interest rate, house loan calculator, soufun.com, decoration. In order to keep consistent with the house price, first convert the weekly Baidu Index into monthly data, and then standardize the data according to the following principles.

$$SS_{it} = \frac{X_{it}}{\sum_{t=1}^n X_i} \quad (1)$$

Among them, SS_{it} represents the Baidu Index of the i keyword in the t period after processing; X_{it} represents the Baidu Index of the i keyword in the t period; n represents the number of months in the selected period. After getting the standardized Baidu Index, calculate the correlation coefficient between each key word and house price, select the key words greater than 0.5 to construct the search index, which are house price SS_1 ($\beta = 0.61$), house tax SS_8 ($\beta = 0.51$) and house loan calculator SS_{12} ($\beta = 0.75$).

Table1. Construction process of search index

Search index	principle	Calculation formula	Correlation coefficient with house price
SS_{15}	Construct by arithmetic average of key words	$0.33 * SS_1 + 0.33 * SS_8 + 0.33 * SS_{12}$	0.74
SS_{16}	Construct with correlation coefficient as weight	$0.33 * SS_1 + 0.27 * SS_8 + 0.4 * SS_{12}$	0.76

As shown in Table 1, in this paper, two methods are used to construct the search index, the first is simple arithmetic average method, and the second is weighted average method based on correlation coefficient. The correlation analysis between the search index and house price shows that the correlation coefficient of the former is 0.74, and the correlation coefficient of the latter is 0.76. Because the correlation between the weighted search index and house price is higher, this paper chooses SS_{16} as the search index.

3.2 Model settings

In the above, we have theoretically analyzed the possible relationship between Internet search and real estate price, and here we will make empirical verification. Because the selected variables are all non-stationary, so we can not directly use the classical linear regression model, so this paper uses the Engle-Granger cointegration test model. If the results of the model show that there is a cointegration relationship, it means that there is a long-term equilibrium relationship between variables, then we can use its correlation to predict. The co integration test is carried out for real estate price (P_1) and search index (SS_{16}). The specific process is as follows:

$$\hat{P}_1 = \hat{c} + \hat{a}_1 SS_{16} \quad (2)$$

$$e_t = P_1 - \hat{P}_1 \quad (3)$$

First, the linear regression equation between variables is established (see (2)), and then the dispersion Series e_t between the real value and the predicted value of the explained variable is obtained (see (3)). At this time, the unit root test is performed on e_t . if the result is stable, there is a cointegration relationship between variables; otherwise, there is no cointegration relationship between them.

4 Empirical results and analysis

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style

your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1 ADF test

Carry out unit root test on variable house price P_1 and search index SS_{16} , and get the results in Table 2. The original time series can not reject the original hypothesis at 1% significant level, which shows that the original sequence is not stable. The first-order difference Series $d(P_1)$ and $d(SS_{16})$ test values are significantly lower than the critical value (- 2.61) of 1%, which shows that variables P_1 and SS_{16} are first-order single integer series.

Table2. Unit root test results of variables

variable	ADF Inspection value	critical value (1%)	conclusion
P_1	0.86	-2.61	Unsteady
$d(P_1)$	-7.49	-2.61	Steady
SS_{16}	1.15	-2.61	Unsteady
$d(SS_{16})$	-2.68	-2.61	Steady

4.2 Cointegration test

Table3. Engle-Granger test results

Cointegration Test-Engle-Granger		
Specification: P_1 SS_{16} C		
Null hypothesis:Series are not cointegrated		
Eigenvalue category	Numerical results	Probability (%)
Engle-Granger tau-statistic	-3.46	4.79
Engle-Granger z-statistic	-18.91	4.64

As both house price P_1 and search index SS_{16} are first-order single integer sequences, in order to determine whether there is a linear relationship, we use the EvIEWS 10 software to carry out the Engle Granger test, and the results are shown in the figure above.

In Table 3, the value of tau statistic is -3.46, and the probability is 4.79%; the value of Z statistic is -18.91, and the corresponding probability is 4.64%. It shows that the two statistics reject the original hypothesis at the significant level of 5%, that is, there is a cointegration relationship between them.

Since there is a cointegration relationship between house price P_1 and search index SS_{16} , it shows that they

have a long-term equilibrium relationship, then we can use search index to build house price model. In order to illustrate the relationship between them, the following regression model is used.

$$P_1 = 24.64 + \frac{0.76}{(8.81)^{***}} * SS_{16} + \varepsilon_t \quad (4)$$

Among them, the adjusted $R^2 = 0.57$, $S.E = 13.18$, $F = 70$, $DW = 0.59$.

According to the results of model (4), it shows that the search index has a positive impact on house prices, and the impact degree is high, the coefficient is 0.76. It means that for every unit change in the search index, the price will change 0.76 units. And because the correlation coefficient is greater than zero, it means that when the search index goes up, the house price will go up, and when the search index goes down, the house price will go down. Therefore, we can use the change of search index to predict the change of house price in the future.

5 Conclusion and Enlightenment

Through the previous analysis, this paper mainly obtains the following conclusions.

Big data has strong prediction ability for house prices in small and medium-sized cities. Through Engle-Granger test of house price index and search index, it is found that there is a co integration relationship between them, and the correlation coefficient is as high as 0.76, which shows that search index has strong prediction ability for Xiangyang house price market. To a certain extent, it also shows that big data plays an important role in regional economic prediction and decision-making. Through reviewing the existing literature, it is found that most of the studies on regional economy are concentrated in large and medium-sized cities, because the economic statistics of small and medium-sized cities are difficult to obtain, which hinders the possibility of economic prediction for small and medium-sized cities, and affects the formulation and implementation of economic decision-making. With the emergence of big data, it has changed the current situation that data in small and medium-sized cities are difficult to obtain, and can provide necessary data support for prediction and decision-making. Therefore, it is of great value to actively explore the impact of big data on the regional economy of small and medium-sized cities. This paper only takes Xiangyang real estate market as one aspect of big data application, hoping to fully tap the potential of big data.

This paper has built a search index, with the correlation coefficient as the weight to perform better. Big data is characterized by a large number of unstructured and jumbled information. Because Baidu search engine is a common browser for Chinese people, Baidu Index provides the public with the function of free query of individual keyword network search volume. Therefore, this paper selects Baidu Index to build the search index. However, there are many key words related to house price. In this paper, key words with correlation coefficient greater than 0.5 are selected for index construction, so as to filter the information of big data to a certain extent. Then, in the process of index construction, we compare the

arithmetic average and the weighted average with the correlation coefficient as the weight respectively. The results show that the search index obtained by the weighted average method has a greater correlation with the house price. Therefore, this paper suggests that we should pay attention to the weight impact of correlation coefficient when using big data to build search index.

Acknowledgment

This work was supported by the grants from Youth Foundation Wuhan Donghu University (2018dhsk007).

References

1. Jiang Zhixiong. "Influence of 'big data + artificial intelligence' technology on economic analysis and prediction," *Knowledge economy*, vol.34, pp.6-13, April 2019
2. Cai Yuezhou. "Big data changing economic forecasting paradigm," *Journal of China Social Sciences*, vol.11, pp.35-42, August 2019
3. Cheng Haixing, Zhu Manzhou. "Research on mechanism and effect of dynamic factor model prediction under the background of big data," *Investment research*, vol.34, pp.142-150, Jan 2015
4. Yang Xi, fan Yingting, Yin Wenjie. "Research on the application of technology transformation investment prediction model based on big data in power industry," *East China electric power*, vol.42, pp.202-206, Feb 2014