

# Review on Second-Hand House Price Forecast

Puxiu Yue, Ruizhen Bai, Yi Wang

School of Information, Beijing Wuzi University, Beijing, China

## ABSTRACT

As a country with a large population, China has a huge real estate market and consumption potential. Nowadays, due to various factors, the second-hand real estate industry began to occupy a large part of the real estate market. Although people buy houses for different reasons, but the price of second-hand housing is the common concern of consumers and developers. Forecasting the price of second-hand housing can not only provide a scientific basis for real estate developers to develop real estate and ordinary residents to buy houses, but also provide a reference for the government to formulate macro-control policies. Therefore, this paper combs the prediction model of second-hand house price and the purpose of purchasing house, hoping to do some research in this area.

**KEYWORDS:** House price forecast; Purpose of house purchase; Multi-linear regression model; Random forest model

**How to cite this paper:** Puxiu Yue | Ruizhen Bai | Yi Wang "Review on Second-Hand House Price Forecast" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-1, December 2020, pp.156-158, URL: www.ijtsrd.com/papers/ijtsrd35852.pdf



IJTSRD35852

Copyright © 2020 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



## 1. INTRODUCTION

### 1.1. Second-hand housing market background

With the continuous growth of China's economy and the continuous development of society, the second-hand housing industry has gradually occupied a large part of the market. The second-hand house prices are rising, and its transaction volume is growing continuously. In the past decade, China's real estate sales revenue has developed rapidly with an average annual growth rate of 2%. Take Beijing as an example, the sales in the central urban area of Beijing maintained a steady development. According to the statistical data, the turnover of second-hand houses in the central urban areas such as Xindongcheng District (the combined urban area of Dongcheng District and Chongwen District, hereinafter referred to as Dongcheng), Xinxicheng District (the combined urban area of Xicheng and Xuanwu, hereinafter referred to as Xicheng) accounted for 8.1% of the total turnover. Haidian and Chaoyang are the most important hot spots, accounting for 25.2% and 18.8% of the total volume respectively (from one day's data). The trading activity of these two regions mainly depends on their huge customer stock and customer demand. Haidian District has the top technology park and talent education base, which gathers a large number of excellent enterprises and high-end talents. Chaoyang District is the location of many business areas and has good facilities and services to attract a large number of people.

According to the survey, the second-hand house has some unique advantages compared with the first-hand house, and the most prominent feature is that the price is more affordable. In addition, on the one hand, residents choose to buy second-hand houses because of their children's schooling, investment value, small for large, easy

employment and suitable for pension. On the other hand, some people choose second-hand houses because the second-hand houses do not need to be decorated, so they can carry their bags directly to save the cost, time and effort of decoration.

### 1.2. Research purpose and significance

Housing is a necessity for the life of an ordinary family, but the reasons for buying a house are not the same. Any change of influencing factors will cause the change of second-hand house price. Therefore, it is the premise and important basis to deeply study and select various influencing factors to predict the change of second-hand house price. Before, many scholars have explored the influencing factors of house price from different angles, mostly from the macro level of policy factors, social factors, economic factors, psychological factors and population factors to predict and analyze the real estate transaction price<sup>[1,2,3]</sup>. However, the simple macroeconomic indicators have certain one-sided and time differences. The results obtained only through the analysis of macro factors are only a general trend, which is far from the actual situation, and has little significance for the purchase choice of ordinary people. In recent years, some scholars have also begun to interpret the second-hand house price from the micro level, but most of the research only stays on the price difference between the residential areas, and the accurate prediction of the price of each room type is very few. In this paper, we hope to use scientific methods to select detailed characteristic variables, so as to achieve the accurate prediction of one house one price. In addition, to analyze the purpose of the residents' purchase, subdivide different types of people, explore the impact of different purchase purposes

on people's selection of second-hand housing and the relationship between the price and this has certain theoretical significance for the improvement of the second-hand house price prediction and evaluation system.

In addition, up to now, scholars at home and abroad have adopted many methods and constructed many models to predict the second-hand house price, such as Classification Regression Tree model, Support Vector Machine (SVM) model, Bagging model, XGBoost model, Lasso model and BP neural network model. Among them, the multi-linear regression model and random forest model are widely used, so this paper will briefly introduce these two models.

The study of second-hand housing price prediction can not only provide a scientific basis for real estate developers to develop real estate and ordinary residents to purchase houses, but also ensure that both parties in real estate transactions can effectively promote their business. At the same time, it can provide a certain reference for the government to formulate macro-control policies.

## 2. Literature review

### 2.1. Overview of second-hand housing price forecast

The multi-linear regression model is the first mathematical model applied to real estate price prediction. Through analysis, the linear relationship between the independent variable and the dependent variable can be found, and the mathematical expression between them can be determined, and the price can be calculated accordingly.

After continuous research and exploration, it is found that there is not a complete linear relationship between the real estate price and its influencing factors, and some qualitative indicators can not be quantified, so scholars began to focus on the construction of nonlinear model. Scholars found that neural network model has the ability to deal with nonlinear problems and has a strong self-adaptive, self-learning ability, which makes it has a unique advantage in real estate price forecasting [4]. Yinglan Qin [5] and others respectively constructed linear regression, regularized regression and artificial neural network (ANN) models to predict the house prices in Japan, and found that the neural network model is better than the regression model. Yuanyuan Li [6] built a BP neural network model to predict the second-hand house prices in Beijing, and the goodness of fit reached 97%. Fei He [7] also established a three-layer BP neural network model to predict the price of second-hand housing in Shanghai in the next quarter.

However, the neural network also has some limitations, it will not work when the data is not enough and prone to over fitting phenomenon. In addition, the objective function it needs to optimize is very complex, which leads to the large amount of calculation and slow convergence speed of neural network algorithm. Then scholars have found a lot of methods to optimize this problem, such as random forest [8,9] and support vector machine. In order to find an optimal method with the highest accuracy, scholars have carried out a lot of experiments. In order to predict the second-hand house prices of six districts in Beijing, Xiaotong Li [10] and others constructed random forest model, SVM model and neural network model, and compared their effects. It was found that the prediction effect of SVM model was second

only to that of random forest model, while the prediction error of neural network model was relatively large. Yijia Chen [11] also built a random forest model to predict the second-hand house prices in Beijing, and used the method of 50% cross test to compare random forest with linear regression, Bagging, simple regression tree, neural network and support vector regression (SVR), and found that random forest has the advantages of small prediction error and high model stability.

### 2.2. Overview of purpose of house purchase

According to the literature and materials, this paper lists seven different characteristics of the purpose of house purchase, covering the vast majority of the current consumer purchase purposes, respectively for a new place to live or for a better house, to buy a house for marriage, to buy a house for living alone, to invest in a house, to work to change a house, to buy a school district house for children's school, and to buy a pension room. Due to the different purposes of purchasing houses, the factors influencing consumers' decision to buy houses are also different. such as, if consumers buy houses for children to go to school, they will consider whether they are located in the school district and whether there are kindergartens and other school district indexes around them; if they buy houses for the elderly, whether there are elevators, greening rate and orientation are the main factors they consider.

### 2.3. Application of multiple linear regression model

Multiple linear regression model is used to infer the situation of independent variables according to the overall situation of multiple dependent variables when a variable is affected by multiple variables. For example, the consumption expenditure of a family is affected not only by the income of family members, but also by the wealth, price level, deposit interest of financial institutions and other factors. In this case, these factors can be used as dependent variables to deduce the approximate consumption expenditure level of the family.

### 2.4. Development and characteristics of Random Forest Model

Compared with the neural network model which has a history of more than half a century, random forest model is a new machine learning model. Although the prediction accuracy of the neural network model is high, it is very computationally intensive. Until the 1980s, Breiman et al. invented the algorithm of classification tree, through repeated binary data classification or regression, greatly reducing the amount of calculation. In 2001, Breiman combined classification trees into random forests. On the premise of no significant increase in the amount of calculation, the random forest model improves the prediction accuracy and saves the operation time greatly. Random forest regression is different from other regression, it is a nonparametric regression technology which has strong adaptability to complex data, and it can effectively analyze the nonlinear, collinear and interactive data. At the same time, it can also analyze the important role of each independent variable to the dependent variable without giving the mathematical form of the model in advance. Random forest has the advantages of fewer parameters that need to be adjusted, no need to worry about over fitting, fast classification speed, and efficient processing of large sample

data. Based on the above advantages, it is currently known as one of the best algorithms so it has been widely used in many fields such as medicine, management and economics.

**Fund: The study was supported by Beijing Social Science Foundation (18GLB022), Beijing Municipal Education Commission “shipei” Project in 2019, Beijing Wuzi University Major Research Projects(2019XJZD12).**

### Reference

- [1] Jin Tian, Research on real estate price forecast of Nanning City under multi factors [D]. Guangxi: Guangxi University, 2018.
- [2] Ruiqi Wu. Research on spatial heterogeneity and influencing factors of second-hand housing prices in Beijing based on GWR Model [D]. Beijing: University of International Business and Economics, 2018.
- [3] Hanzhong Zhang, Qian Zhang, Bin Wang, et al. Commodity housing price distribution and its influencing factors in Shanghai [J]. China market, 2018 (2): 42-48.
- [4] Yuming Gao, Renjin Zhang. House price prediction and analysis based on genetic algorithm and BP neural network [J]. Computer Engineering, 2014, 40 (4): 187-191.
- [5] Yinglan Qin, Yuki Akiyama, Yoshiki Ogawa, Ryosuke Shibasaki, Taisei Sato. Study on the relationship between house rent and people congestion by time in Tokyo based on mobile phone GPS data [J]. 2019 IEEE Big Data Conference. Dec 2019, Los Angeles.
- [6] Yuan-Yuan Li. Research on transaction price prediction of second -hand house in Beijing based on BP neural network [D]. Beijing: Capital University of Economics and Business, 2017.
- [7] Fei He, Verification and prediction of second-hand house price in Shanghai based on BP neural network [D]. Shanghai: University of Shanghai for Science and Technology, 2018.
- [8] Muxi Yang. Price evaluation of second-hand housing based on Stochastic Forest Model [D]. Hunan: Central South University, 2012.
- [9] Zheng Pang. Study on the application of Stochastic Forest Model in batch evaluation of second-hand housing [D]. Chongqing: Chongqing Jiaotong University, 2017.
- [10] Xiaotong Li, Xuan Guo, Chengjie Wang. Study on the price of second-hand housing in Beijing based on random forest method [D]. Beijing: China University of petroleum, 2017.
- [11] Yijia Chen. Study on the evaluation model of second-hand housing in Beijing based on Stochastic Forest theory [D]. Beijing: Beijing Jiaotong University, 2015.

