

Analysis of Real Estate Market Pricing Rules and Construction of Prediction Model from the Perspective of Big Data

Xinyi Cao

*Qiantang High School, Hangzhou, China
casino17746837986@outlook.com*

Abstract. Amidst the digital revolution, data resources are surging exponentially. Cutting-edge technologies like big data and AI are rapidly evolving, creating new opportunities for industries. Yet, the real estate market is becoming more complex and unstable. Frequent policy changes, supply-demand imbalances, and economic volatility are heightening industry uncertainty. Meanwhile, consumers are demanding better living experiences, smarter amenities, and greener homes. These diverse and refined needs are driving a major shift in the real estate landscape. This paper employs a method of literature review and case study. It investigates the influence of big data in the prediction of price and the construction of the real estate market. This big data analysis on housing prices holds significant value by empowering homebuyers to understand explicit/implicit factors for informed decisions, guiding real estate firms on land/pricing strategies, supporting government policy formulation with data, aiding macroeconomic stability through risk identification in price patterns, and enriching academic theory in real estate economics through systematic analysis and novel methods.

Keywords: Big data, Analysis, Real estate, Market, Pricing

1. Introduction

In the current era of the digital wave sweeping across the globe, the vast amount of data generated daily is several times the total amount of data accumulated from the beginning of human civilization to the year 2003. Big data and artificial intelligence technologies are iterating at a speed of 15% per month. In terms of policy, the core of real estate policies in 2025 is to “boost demand, optimize supply and enhance financial coordination. By implementing measures such as city-specific policies to reduce restrictions, tax and fee reductions, credit expansion, increasing the supply of affordable housing, urban village renovation, stock acquisition, and promoting good houses, the market is driven towards a stabilization of both volume and price, achieving a stop-loss and recovery. However, the market remains in an “L-shaped” bottoming-out phase, with a prominent structural contradiction between the partial recovery of core cities and the in-depth adjustment of small and medium-sized cities. On the supply and demand side, the coexistence of population aging and

unbalanced urbanization has led to a sharp contrast between the overstock in third-and fourth-tier cities and the structural shortage in first-tier cities.

This paper selects the housing sales data in Ames, Iowa, USA, from 2006 to 2010 to build a model for estimating the final selling price of houses. Due to the large number of explanatory variables, models such as linear regression, random forest, and gradient boosting model are used for feature selection to identify the significant features affecting the final selling price of houses. The research shows that the gradient boosting model performs the best and the results are the most effective [1-2]. This paper applies spatial data analysis and modeling techniques to achieve three research objectives: analyze the spatial distribution pattern of housing prices at the micro level, explore the spatiotemporal evolution characteristics of residential properties and detect the geographical differences in housing prices. The analysis results clearly present the spatiotemporal patterns of housing prices in metropolitan areas, fully verifying the application value of big data and its analysis methods [3].

Big data, through its cross-analysis of massive amounts of data, including geographic information, credit records and social media sentiment, has become the key to solving industry challenges and is driving the real estate market from an experience-driven model to a data-smart era. With big data analytics, real estate companies can anticipate market turning points, optimize their investment layouts, develop differentiated pricing strategies, and precisely meet the needs of different consumer groups, thereby maximizing profits. In conclusion, the research in this paper not only provides practical guidance for various market participants at the practical level but also has certain academic value at the theoretical level. It is of great significance for promoting the stable and healthy development of the real estate market and the stability of the macroeconomy.

2. Factors affecting housing price changes

The main factors that influence housing prices include geographical location, area size, and surrounding facilities, among others [4-5].

At the city level, economically developed cities, such as first-tier cities and some second-tier cities, typically offer more job opportunities, business prospects, and better infrastructure. As can be seen from Figure 1, these factors attract a large influx of population, increase housing demand, and thus drive up housing prices. For example, Shenzhen, as one of China's technology and financial centers, has experienced rapid economic development. It has attracted numerous high-tech companies and financial institutions, leading to a strong demand for housing due to the large employed population, and consequently, relatively higher housing prices. And at the regional level, the convenience of transportation and the availability of educational resources are both significant influencing factors.

At the micro level, the topography and geological conditions of the land also impact housing prices. Land that is flat and geologically stable has relatively lower development costs and more stable housing prices. However, if the land is prone to geological disasters, such as being located in an earthquake zone or an area susceptible to landslides and mudflows, the risk associated with the property is higher. Housing prices may be suppressed, or developers may need to implement more safety measures during development, which increases costs and also affects housing prices. Regarding the area size of houses, I have collected data on housing prices for different sizes and plotted them on a scatter chart. The chart shows a positive correlation between area size and housing prices, but the correlation is not very strong. This is because first-time homebuyers, who have limited budgets, are more concerned with the "low total price threshold." To meet this demand, developers design small apartments as high-cost-performance products. They increase the unit price

while reducing the area to lower the total price. Additionally, small apartments usually have higher rental return rates than larger ones (such as school district houses or houses near subway stations), which attracts investors and drives up their unit prices.

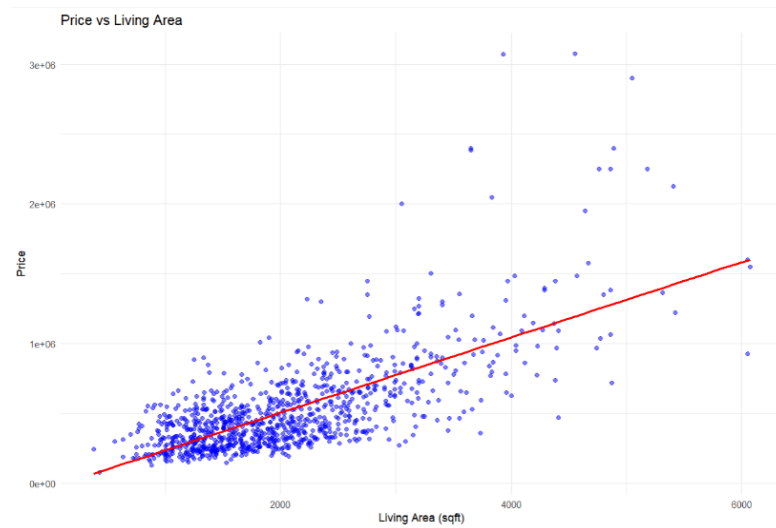


Figure 1. Price VS using area

3. The hidden patterns of housing prices

The hidden patterns behind housing price increases are mainly related to economic and monetary policies, cost-push factors, psychological expectations, and inflation. From the perspective of big data, the first point is that when economic growth is rapid and the money supply increases, there is an abundance of funds in the market. This enhances the purchasing power of homebuyers, and housing prices tend to rise. Conversely, during economic recessions or when monetary policies are tightened, housing prices may be suppressed. For example, following the 2008 global financial crisis, housing prices in many countries and regions either fell or stagnated. However, as the economy recovered and monetary policies became more accommodative, housing prices gradually rebounded.

The second point is that big data can reflect the impact of building material prices, land costs, and labor costs on housing prices. Although the proportion of costs such as building materials in the composition of housing prices is limited, when these costs rise, developers will pass on a portion of the increased costs to housing prices. Land costs are an important component of housing prices. By analyzing the relationship between land auction prices and housing prices, it can be found that in areas with high land auction prices, the cost of housing development is high, and housing prices are correspondingly higher. For example, in some city center locations where land is scarce and land auction prices continue to rise, the prices of new housing projects are relatively high.

The third point is that the psychological expectations of homebuyers have an impact on housing prices, which can be reflected through data such as the volume of housing transactions and listing prices. When the market generally expects housing prices to rise, homebuyers enter the market earlier, increasing demand and driving up housing prices. When the expectation is that housing prices will fall, homebuyers tend to hold off on purchases, reducing demand and causing housing prices to fall or slowdown in growth. For example, when the media reports an upward trend in housing prices or significant positive news about a particular area, the volume of housing transactions in that area may increase, and housing prices may rise. Conversely, when there is

negative news or economic instability, homebuyers become more cautious, transactions shrink, and housing prices are suppressed.

The fourth point is that big data shows that when the price index of goods rises, especially the prices of agricultural products and energy, housing prices usually increase. This is because inflation devalues currency, and real estate, as a tangible asset, has the function of preserving and increasing value. Investors and homebuyers are more inclined to invest in real estate to hedge against inflation risks. By analyzing housing price data under different levels of inflation, it can be found that there is a certain positive correlation between the inflation rate and housing prices. During periods of high inflation, the increase in housing prices is often more significant.

4. House price calculation method

Using historical data to predict housing prices is also a good approach [6-7]. For example, it can capture market patterns, reduce decision-making blindness, provide data support for policy formulation, help businesses optimize their investment and operational strategies, and so on.

Firstly, through the time series data in historical records (such as the quarterly increases in housing prices over the past 10 years), it is possible to identify market cycle phases (such as boom periods or adjustment periods). This helps avoid blindly entering the market at a peak or panicking and selling at a trough. If historical data shows that housing prices in a certain area have been rising by an average of 3-5% annually without any signs of reversal, it is reasonable to expect future trends and provide a basis for long-term asset allocation (such as calculating the returns on pension fund investments in real estate).

Secondly, the government can use historical data to simulate the outcomes of policy interventions. For example, by referring to historical data showing that housing prices fell by 1-2% month-on-month after a “tightening of purchase restrictions” in the past, it can assess the effectiveness of the current tightened purchase restrictions in curbing the rapid rise in housing prices. Based on the historical pattern that an increase of 10% in the supply of affordable housing led to a 3% drop in the prices of surrounding commercial housing, the government can set a reasonable scale for the construction of affordable housing to avoid impacting the commercial housing market.

Thirdly, from the perspective of land investment, historical data can be used to compare the housing price growth potential of different areas and guide land acquisition decisions, avoiding the purchase of land at high prices in low-growth areas. In terms of pricing strategies, the pricing of new housing projects can refer to the historical transaction price trends of surrounding second-hand houses.

In summary, using historical data to predict housing prices has core advantages such as a scientific basis, quantifiability, and low cost. It can provide decision-making support for multiple stakeholders, including governments, businesses, and investors.

5. Conclusion

In summary, the main body of the discussion focuses on how housing prices are influenced by a multidimensional array of factors, including geographical location, housing size, economic and monetary policies, costs, psychological expectations, and inflation. In terms of geographical location, the level of urban economic development, regional transportation and educational resources, as well as land and geological conditions, all have an impact on housing prices. Housing size has a weak positive correlation with housing prices, which is influenced by both rigid demand

and investment demand. Factors such as economic growth, loose monetary policy, rising costs, and inflation expectations tend to drive up housing prices, while the opposite tends to suppress them. Historical data, through time series analysis and policy simulation, can capture market patterns and provide a scientific basis for decision-making for government regulation, corporate investment, and individual home purchases, reducing the blindness of such decisions. The future trend of housing prices needs to be continuously tracked by integrating macroeconomic conditions, policy dynamics, and regional development trends, with attention to changes in long-term factors such as population movement and land supply. However, the study still has some limitations. For example, it did not analyze historical sudden events and policy changes. In the future, the author will investigate deeper in this field with other useful analysis tools.

References

- [1] Singh, A., Sharma, A., & Dubey, G. (2020). Big data analytics predicting real estate prices. *International Journal of System Assurance Engineering and Management*, 11, 208-219.
- [2] Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, 11(3), 334.
- [3] Li, S., Ye, X., Lee, J., Gong, J., & Qin, C. (2017). Spatiotemporal analysis of housing prices in China: A big data perspective. *Applied Spatial Analysis and Policy*, 10(3), 421-433.
- [4] He, H. M., Chen, Y., Xiao, J. Y., Chen, X. Q., & Lee, Z. J. (2021). Data analysis on the influencing factors of the real estate price. *Artificial Intelligence Evolution*, 52-66.
- [5] Trindade Neves, F., Aparicio, M., & de Castro Neto, M. (2024). The impacts of open data and eXplainable AI on real estate price predictions in smart cities. *Applied Sciences*, 14(5), 2209.
- [6] Xiao, Y. (2022). Big data for comprehensive analysis of real estate market.
- [7] Yuan, Y. (2025). Research on Real Estate Market Prediction Models Based on Big Data and Artificial Intelligence. *International Journal of High Speed Electronics and Systems*, 2540439.