ISSN 2959-6149

Analyzing Influencing Factors and Inter-city Differences of House Prices Based on Statistical Methods

Yingqi Xu*

Department of Brunel College, North China University of Technology, Beijing, China
*Corresponding author: 21190010104@mail.ncut.edu.cn

Abstract:

With the acceleration of urbanization and the improvement of residents' living standards, house prices have become a key factor affecting the quality of life of urban residents. Differences in house prices between cities can be attributed to a variety of factors, including, but not limited to, the level of economic development and residents' incomes. To gain a deeper understanding of the variability of house prices and its determinants across cities, this study selects five representative cities in the UK for analysis. It explores whether there are statistically significant differences in house prices through Analysis of Variance and applies multiple linear regression modeling to assess the specific roles of each influencing factor on house prices. This analysis aims to provide policymakers and market participants with an in-depth understanding of the trends in house price changes, thereby supporting the effective allocation of real estate resources and the balanced development of the economy. The results not only reveal significant differences in house prices across cities but also identify key factors that significantly impact house prices, providing insights to promote the healthy development of the real estate market.

Keywords: Box-and-Whisker Plot; Analysis of Variance; multiple linear regression modeling.

1. Introduction

The rigid demand for housing, an important part of people's consumption, has driven the continued prosperity of the property industry. As a result, the issue of housing prices has been a focus of great attention for the financial industry and investors. In a market economy, house prices not only affect the cost of living of residents, but also are a key factor in urban development and social stability. Therefore, domestic and foreign scholars generally regard the study of house prices as an important research field. With the advent of the big data era, the availability of public data has significantly increased, providing rich data resources for house price research. These data cover information on house prices, residents' income levels, urban planning and infrastructure development in various cities, providing an empirical basis for analyzing and predicting house prices. By analyzing these data in depth, scholars are able to explore the variations in home values among various areas and further investigate the economic, social, policy and geographical factors that contribute to these differences.

Using data from 1988 to 2012, Elias and Janne argued that regional differences in house prices in the Finnish market are generally quite small, and that the long-term coefficients of income vary by region, with cities having stronger supply constraints exhibiting larger long-term coefficients [1]. China's house prices show a regional declining tendency from east to west, according to a research by Yi and Liu based on panel data for 31 provinces from 2013 to 2018, influenced by factors such as income, population, and household savings, with a greater impact noted in the eastern regions [2]. Chris and Jonathan, using price data from 1969 to 2016, tracked several economic cycles and noted the widening gap between London, particularly central London, and the rest of the UK. They argued that although Brexit may have ended London's recent long-term boom, there is no indication that regional house price gaps will narrow [3]. In 1989, Mankiw and Weil concluded that house prices were rising due to population growth, marking the first study to link population growth directly with rising house prices [4]. In 1996, PotePan utilized the characteristic price model, employing variables such as population, interest rates, income, building expenses, and public services quality to analyze the factors influencing along with real estate values in 58 US urban areas between 1974 and 1983. The results indicated that income and construction costs were the most important factors affecting inter-regional real estate prices [5]. Geoff Kenny analyzed the supply and demand of real estate in Ireland by introducing long- and short-term variables, finding that

income levels and construction costs positively affect prices, whereas mortgage interest rates have a negative impact [6]. Miller and Peng examined the potential time-varying volatility of single-family home value appreciation using a panel VAR model and a GARCH model. Their findings indicated that future home prices and per capita income were important factors contributing to house price volatility [7]. In their analysis of the macroeconomic effects on housing prices, Zeno and Roland used thirty years' worth of data from fifteen different nations. They found that, on average, a rise in building costs would have a long-term impact of 0.6% on housing prices. This increase leads to higher rents, which subsequently drive up house prices [8]. In 2020, Morteza and Daniel examined how immigration affected home values in 2006, 2011, and 2016. They discovered that while immigration raises home values, the annual decrease in Australian house prices would be approximately 1.1% in the absence of migrants [9]. Braun and Lee estimated changes in residential land values and prices in 379 German counties between 2014 and 2018 by analyzing data from 42,647 observations. Their research revealed that land values rose by 27% on average, home values by 20%, and that cyclical fluctuations in land values in Germany are probably going to have an impact on how house prices develop going forward [10].

By reading the literature, it was learned that regional differences in house prices and the factors affecting them constitute a complex global topic, involving a broad array of economic, social, policy, and geographical factors. Despite the plethora of international studies, these predominantly focus on the macroeconomic level and cross-regional comparisons, rarely delving into the specific administrative regions of a given city. To more accurately analyze the mechanisms of house price formation and intra-regional nuances, this study focuses on five London boroughs: Harrow, Barnet, Enfield, Brent, and Hillingdon. By analyzing these specific areas in depth, the study will first determine whether there are significant differences in house prices among these areas through analysis of variance, and then further explore the area-specific factors influencing these differences. Additionally, one of the administrative districts will be selected for in-depth analysis to identify the key factors affecting house prices in that region through multiple linear regression modeling. This study aims to provide an empirical basis for local policymakers to support them in making more precise decisions in property market management and urban planning. Simultaneously, this study will also furnish the field of house price research with a case study on how to analyze the factors influencing house prices within a city, thus enriching the current theory and practice of house price research.

2. Methods

2.1 Data

The dataset used in this study is derived mainly from official statistics released by the UK government, which cover house prices and related economic indicators for each London borough as of the year 2016. To comprehensively analyze the mechanisms of house price formation, five representative boroughs in London are selected: Harrow, Barnet, Enfield, Brent, and Hillingdon. Furthermore, to explore the factors behind the high house prices, detailed data on economic indicators such as household income and employment rates are specifically collected for the borough with the highest house prices.

2.2 Methodology

Analysis of variance (ANOVA) is a statistical method used to test whether the means of three or more groups of data are significantly different. By analyzing and comparing the variance within different groups and the variance between groups, ANOVA helps us to determine whether the groups of data are likely to come from an overall population with the same means to determine whether there is a statistically significant difference [11].

Multiple linear regression analysis establishes models to determine quantitative relationships between two or more interdependent variables. It uses the least squares method for parameter estimation [12].

3. Result and Discussion

3.1 Data Distribution

This paper provides a visual comparison of the data through box-and-line plots. By analysing the median, interquartile range, outliers, etc. a preliminary understanding of house price data for different cities is provided.

Fig. 1 Box-and-Whisker Plot of House Prices in Five Cities

Figure 1 reflects the fact that the median seems to be around £400,000 for all cities. However, Barnet‘s median is slightly higher than the others, which could mean that house prices are generally higher in Barnet than in the other four cities. The length of each box shows the range from the first quartile (Q1) to the third quartile (Q3), i.e., representing the middle 50 per cent of the distribution of the data. The narrower quartile range for Barnet suggests that house prices are more concentrated. By contrast, the wider quartile ranges for Harrow and Brent suggest that house prices are more widely spread in these cities. Harrow and Brent have very long whiskers in the chart, indicating a wider range of house price fluctuations in these cities.

3.2 Problem analysis

In this study, we use one-way analysis of variance (ANOVA) to explore the statistical differences in house prices between cities.

Null Hypothesis (): Equal population averages for the five cities. One way to phrase the hypothesis is:

Alternative Hypothesis ():

Sum of squares for total (SST):

(1)

It is the sum of the squares of the errors of all observations and the total mean .

Sum of Suares for factor A (SSA):

(2)

It is the sum of the squared errors of the group means and the total mean .

Sum of squares for error (SSE):

(3)

It is the sum of the squares of the errors between each sample’s data and the mean of its group at each level or group.

Next, the significance of the differences between groups was determined by comparing the within-group variance to the between-group variance.

(4)

(5)

(6)

```
Total Sum of Squares (SST): 12125534782 68.842285
Between-group Sum of Squares (SSA): 217195507254.353210
Within-group Sum of Squares (SSE): 995357971014.489014
F Value: 5.782524
p Value: 0.00030038269452725697
```

Fig. 2 ANOVA calculations

The results in Figure 2 show that the F-statistic for the between-group ANOVA is 5.78254, with a corresponding p-value of 0.0003. This p-value is significantly smaller than the commonly used significance level of 0.05, thus strongly rejecting the null hypothesis that the house prices in the five cities are equal. This suggests that there is a statistically significant difference in house prices between at least two cities. This statistical significance suggests that the variation in house prices between cities is beyond the range of random errors. It can therefore be concluded that city characteristics have a significant impact on home price differences.

After conducting a one-way ANOVA to investigate the variability of house prices across five cities, Significant differences among them were identified. Subsequently, the paper honed in on the city with the highest house prices for a more detailed examination. To decipher the pivotal factors influencing the housing market in this specific urban setting, a multiple linear regression model was employed. House prices were posited as the dependent variable, with population density, regional area, employment rate, and median household income serving as independent variables. This analysis facilitated the quantification of the independent contribution of each variable to house prices and enabled an assessment of their relative significance. The aim of this analysis is to reveal the principal determinants of house prices within this city and to evaluate the extent of their impact.

The linear regression model is expressed as:

(7)

```
Regression coefficient:
803502.685789
-1.590828
997.557293
23291.371161
32.400075
Multiple coefficients of determination R^2: 0.890137
```

Fig. 3 Linear Regression Equation

As illustrated in Figure 3, the first row is a constant term that represents the base house price when the influences included in the model are not considered. The following rows show the correlation coefficients between population density, area size, employment rate, and household income and house prices, respectively. According to the results of the model, both area size and employment rate have a significant favorable impact on house prices, which could result from the fact that larger homes or parcels imply the possession of more scarce resources, as well as the fact that house prices are generally higher in areas with high economic and income levels. Household income is also positively correlated with house prices, with areas that tend to have higher income levels being able to support higher house prices. In contrast, the slight negative impact of population density on house prices may reveal the negative effects of congestion and competition for resources on the attractiveness of living environments brought about by high population densities.

Next the residuals are plotted to assess how well the model fits the factors influencing house prices.

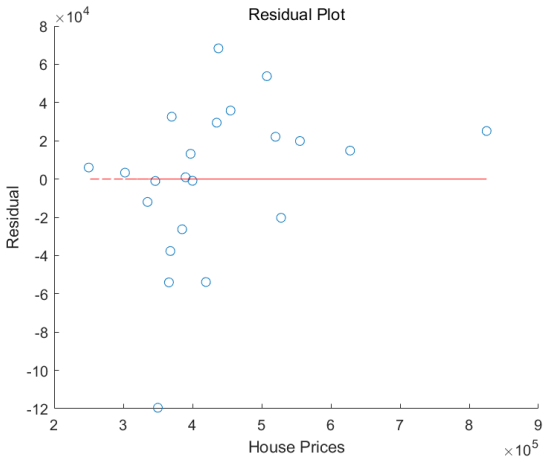

Fig. 4 Residual Plot

In the residual plots presented in Figure 4, most of the residuals are closely distributed around the zero line and the distributions do not exhibit any systematic bias. This pattern suggests that the linear regression model employed is effective in capturing the inherent variability of the data over a wide range. However, some residual values can also be seen that deviate significantly from the zero line, and these significant deviations may point to unique attributes of a particular property, such as the uniqueness of its geographic location and its potential historical value, which may not have been adequately accounted for in the current model.

4. Conclusion

This study provides insights into the mechanisms of house price formation and the intra-regional differences by analyzing house prices and their associated economic indicators in detail across five London boroughs. Through Analysis of Variance, this paper confirms the presence of significant differences in house prices among these boroughs and employs multiple linear regression modeling to further elucidate the key economic factors driving these variations. Although this study is based on authoritative official statistics, it encounters certain limitations, primarily that the dataset may not comprehensively capture all potential factors affecting house prices, such as short-term economic fluctuations and unexpected social events. Moreover, while ANOVA and multiple regression modeling lay a foundation for understanding movements in house prices, these methods might not adequately address complex non-linear relationships. Consequently, future research should consider adopting more advanced statistical methods or machine learning techniques to enhance the models' predictive accuracy and their capacity to manage complex data structures. Employing these methods, we anticipate delivering a more detailed depiction of house price trends, thereby better facilitating policy development and promoting the healthy growth of the real estate market.

References

[1]Oikarinen, E., & Engblom, J. (2016). Differences in housing price dynamics across cities: A comparison of different panel model specifications. Urban Studies, 53(11), 2312-2329.

[2]Yi, Y., & Liu, M. (2019). Research on regional differences in the influencing factors of China's house prices. Times Finance, (29), 45-46, 57.

[3]Hamnett, C., & Reades, J. (2019). Mind the gap: Implications of overseas investment for regional house price divergence in Britain. Housing Studies, 34(3), 388-406.

[4]Mankiw, N. G., & Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. Regional Science and Urban Economics, 19(2), 235-258.

[5]Potepan, M. J. (1996). Explaining intermetropolitan variation in housing price, rents and land prices. Real Estate Economics, 24, 219-245.

[6]Kenny, G. (1999). Modelling the demand and supply side of the housing market: Evidence from Ireland. Economic Modeling, 16, 389-409.

[7]Miller, N. G., & Peng, L. (2006). Exploring metropolitan housing price volatility. The Journal of Real Estate Finance and Economics, 33(1), 5-18.

[8]Zeno, A., & Roland, F. (2009). Macroeconomic determinants of international housing markets. Journal of Housing Economics, 19(1), 1-13.

[9]Morteza, M., & Daniel, M. (2020). The impact of immigration on housing prices in Australia. Papers in Regional Science, 99(3), 773-786.

[10]Braun, S., & Lee, G. S. (2021). The prices of residential land in German counties. Regional Science and Urban Economics, 89, 22-29.

[11]Long, H. M., & Guo, W. (2009). Dynamic econometric analysis of house price and land price in China based on VAR model. Economic Mathematics, (2), 52-58.

[12]Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019, March). House price prediction using regression techniques: A comparative study. In 2019 International conference on smart structures and systems (ICSSS) (pp. 1-5). IEEE.