The Simpson's Paradox of the Proportion of School District Housing and Housing Prices: A Comparative Analysis of Hai'an City and Jiangyan District

Zihan Xu*

School of Statistics, East China Normal University, Shanghai, 20062, China

*760541220@qq.com

Abstract. In order to explore the existence of Simpson's paradox in housing prices, a case study method is adopted to conduct statistics and analysis on the housing prices in Hai'an City and Jiangyan District, Jiangsu Province. The results show that although the overall housing price in Hai'an City is lower than that in Jiangyan District, the housing prices of both school district housing and non-school district housing in Hai'an City are higher than those in Jiangyan District, which fully proves the existence of Simpson's paradox.

Keywords: Simpson's Paradox; Housing Price; School District Housing

1 Introduction

The overall situation in the current real estate market presents a complex and changeable pattern. Against this backdrop, there are significant differences in the performance between school district housing and non-school district housing, and the influence of their proportional relationship on housing prices has become more prominent. School district housing, due to its close connection with high-quality educational resources, has always been a special existence in the real estate market. Parents are eager to purchase school district housing in order to provide their children with better educational opportunities, making school district housing often in short supply in the market. The price trend of non-school district housing is more affected by conventional factors such as market supply and demand, location, and supporting facilities. In some new urban areas, due to the incomplete infrastructure construction, the price growth of non-school district housing is relatively slow, and there may even be a certain degree of price correction during the market downturn. In some third- and fourth-tier cities, with the outflow of population and the increase in housing inventory, the price pressure on non-school district housing is more obvious.

The proportions of school district housing and non-school district housing vary among different cities, and this difference has a profound impact on housing prices, even leading to the emergence of Simpson's paradox. Simpson's paradox refers to the situation where two sets of data under a certain condition both satisfy a certain property when discussed separately, but once considered together, it may lead to the oppo-

© The Author(s) 2025

B. Siuta-Tokarska et al. (eds.), *Proceedings of the 2025 3rd International Academic Conference on Management Innovation and Economic Development (MIED 2025)*, Advances in Economics, Business and Management Research 348,
https://doi.org/10.2991/978-94-6463-835-6_61

site conclusion. In the real estate market, this phenomenon is manifested as follows: for example, in City A, the proportion of more expensive school district housing is relatively high; in City B, the proportion of school district housing is relatively low. Overall, the total housing price in City A is higher than that in City B, but when the data of school district housing and non-school district housing in City A and City B are analyzed separately, it may be found that both the school district housing and non-school district housing in City A are lower than those in City B. Conducting an in-depth study of this Simpson's paradox in housing prices caused by the different proportions of school district housing and non-school district housing is of great significance for accurately grasping the price trend of the real estate market.

This study will adopt the case study method to deeply analyze the phenomenon of Simpson's paradox in the housing prices of the two cities in 2024 due to the different proportions of school district housing and non-school district housing, and comprehensively reveal the underlying deep-seated reasons. Two representative cities will be selected as the research objects, and the actual situations of school district housing and non-school district housing in their real estate markets will be deeply analyzed. Through detailed case studies of these two cities, first-hand data and materials, including housing price data, housing transaction information, market supply and demand situations, etc., will be obtained, providing a solid data foundation for the subsequent analysis.

2 Theoretical Basis and Literature Review

2.1 Simpson's Paradox Theory

Simpson's paradox (Simpson’s paradox) is a unique and confusing phenomenon in the field of statistics, which can be traced back to the end of the 19th century. In 1899, Karl Pearson, a British statistician, discovered in his research that there may be differences between the marginal correlation and partial correlation of continuous variables, which laid the foundation for the proposal of Simpson's paradox. In 1903, George Udny Yule further pointed out the paradoxical phenomenon existing in the correlation of categorical variables. It was not until 1951 that Edward Simpson formally and systematically explained this phenomenon in his thesis, and this phenomenon was thus named Simpson's paradox, also known as the Yule-Simpson effect (Yule - Simpson effect).

In terms of definition, Simpson's paradox means that when studying the correlation between two variables, the party that has an advantage in the grouped comparison does not have an advantage in the overall evaluation when the variables are studied separately. Mathematically, if there are four variables a, b, c, and d, under the grouped situation, a certain relationship may exist, but when the data are combined, a different relationship (where e, f, g, h are another set of related data) may occur.

Its principle mainly stems from the data grouping method and the distribution characteristics of the data in each group. In actual situations, there may be some unobserved or uncontrolled variables between different groups, and these variables will

interfere with the results, resulting in a reversal of the relationship between the whole and the parts.

2.2 Literature Review

In terms of empirical research, scholars have carried out a large number of case analyses. Hamnett and Bramley (year) studied the school district housing in London and found that the prices of high-quality school district housing were significantly higher than those of ordinary school district housing. They further verified the influence of school quality on housing prices through the collection and analysis of housing price data in different school districts. [1] In Australia, Firmin Doko Tchatoka and Vanessa Varvaris studied the primary school district divisions in New South Wales, Australia, and found that properties close to high-quality schools not only had higher prices but also circulated faster in the market. Their research also shows that with the increasing emphasis on education, the price gap between school district housing and non-school district housing has a tendency to gradually expand. [2] When Sam DeBord investigated the housing prices of school district housing in Seattle, he found that most of the buyers of school district housing near the best schools were Chinese buyers, which indicates that Chinese people often attach great importance to educational resources and pay more attention to school district housing. [3]

In China, research on the relationship between school district housing and housing prices has been increasing in recent years. At the theoretical level, scholars have analyzed it from multiple perspectives such as the supply and demand theory and the location theory. Scholars such as Wang Xiaodong pointed out that "There is an imbalance in educational resources in the basic education stage in China, which has given rise to the phenomenon of'school district housing', linking educational opportunities with family wealth, exacerbating social inequality in education, and also leading to the distortion of the real estate market, with the prices of school district housing rising irrationally." [4]

There are also some deficiencies in current domestic research. The data sources of some studies are relatively single, lacking a comprehensive analysis of multi-source data. There is relatively little research on the influence of the proportion of school district housing and non-school district housing on housing prices, especially on Simpson's paradox. Most studies mainly focus on the direct relationship between school district housing and housing prices, while ignoring the influence of the differences in the proportions of school district housing and non-school district housing between different cities on the overall trend of housing prices.As Wang Xiaodong et al. (2024) demonstrated, this oversight may lead to misinterpretations of housing price dynamics, especially when analyzing segmented markets like school districts [5]. Based on this, this study focuses on the phenomenon of Simpson's paradox in the housing prices of the two cities in 2024 due to the different proportions of school district housing and non-school district housing. Through in-depth analysis of this phenomenon, it provides a new perspective and ideas for the research of the real estate market.

3 Selection of Case Cities and Data Collection

3.1 Case Selection

Hai'an City is a key county-level city under development in Jiangsu Province and is in a stage of rapid urbanization. It has well-known educational institutions such as Hai'an Senior High School in Jiangsu Province, and its educational quality has considerable influence among county-level cities in the whole province and even across the country, forming a distinct concept of school districts. Jiangyan District also has rich educational resources. Jiangyan Middle School in Jiangsu Province has an excellent performance in the national middle school rankings, attracting a large number of students and parents in the local area and even a wider range, which also makes the school district housing market quite active. The advantages of their educational resources and the formation of the school district housing concept provide typical samples for this study.

Hai'an City has a stable economic development and a diversified industrial structure, and the supply and demand relationship in the real estate market is relatively stable. It is neither disturbed by too many complex factors like the real estate market in first-tier cities nor lacks market activity like some small cities. Jiangyan District, as an important area of Taizhou, has a good economic foundation, and its real estate market is also relatively stable, being relatively less affected by macroeconomic fluctuations. Such a moderate market environment makes the research results more representative and valuable for reference when studying the relationship between school district housing and the total housing price of the two cities.

3.2 Data Collection Sources and Methods

3.2.1 Data Source Channels.

This article's data covers real - estate market macro - data, housing details, and transaction data. Sources include government real - estate departments, real - estate platforms like Fang.com and Ke.com, and relevant statistical institutions.

3.2.2 Data Collection Methods and Sample Selection.

In terms of sample selection, in order to ensure the representativeness and reliability of the research results, this study has formulated strict sample selection criteria. For the samples of school district housing, all the communities around all the key schools in Hai'an City and Jiangyan District have been selected. These key schools cover primary schools, junior high schools, and senior high schools to ensure that the educational needs of different age groups can be reflected.

Finally, this study has collected 23 samples of school district housing in Hai'an City and 65 samples of school district housing in Jiangyan District. These samples cover houses in different regions, different price ranges, different house ages, and different layouts, which can comprehensively reflect the price characteristics and market situations of school district housing and non-school district housing in Hai'an City, providing a solid data foundation for the subsequent data analysis and research.

4 The Manifestation of Simpson's Paradox in the Housing Prices of the Two Cities

4.1 Analysis of the Housing Price Data in Hai'an City

The following is a detailed table obtained from the survey:

As shown in Table 1, the selected representative communities in Hai'an City exhibit significant variations in average house prices and household numbers. For instance, Xintuo • Donggang Guoji has an average price of 16,519 CNY, while Haizhou Huayuan is priced at only 7,954 CNY.

Table 1. Selected Representative Communities with Average House Prices and Households

Community Name	Average House Price (CNY)	Number of Households
Xintuo • Donggang Guoji	16,519	1,674
Shuxiangyuan	16,340	1,268
Ruihai • Shengshimingmen	15,867	1,312
Wanda Haizhixin	12,435	2,379
Jinshi Lanjun	12,936	1,502
Sujian Xincun	10,600	409
Zhongyang Xiandai Cheng	10,652	1,378
Huaxin Yipin	9,363	1,909
Jiguanbei Xiaqu	8,920	221
Haizhou Huayuan	7,954	256
The Community at No.20 Renmin West Road	8,500	3

4.2 Analysis of the Housing Price Data in City B

The following is the detailed table obtained according to the survey.

Table 2. Selected Representative Communities with Key Metrics

Community Name	Average House Price (CNY)	Number of Households
Jintong Danguiyuan	18,000	242
Zhongtian Taoyuan Yaju	17,368	397
Runjing Guoji Huayuan	14,756	768
Langyuewan	13,000	1,950
Fenghuangyuan	12,683	1,113
Wangjiashe Xiaqu	12,100	279
Jinxiu Jiangcheng	10,553	3,780
Jun'an Shijicheng	10,100	3,880
Futian Mingyuan	10,881	894
Huayang Wenjinyuan	9,722	2,177
Xinshiji Huayuan	8,975	520
Rongjin Huayuan	8,231	1,083

In Table 2, the representative communities in Jiangyan District display a similar trend, with Jintong Danguiyuan reaching 18,000 CNY, while Rongjin Huayuan is priced at 8,231 CNY.

4.3 Comparative Analysis of the Simpson's Paradox in the Housing Prices of the Two Cities

According to the relevant data from the official websites of the People's Government of Hai'an City and the People's Government of Jiangyan District, the average price of houses that are not in school districts can be calculated. The following is the sorted-out data. Table 3 provides a clear comparison of house prices between Hai'an City and Jiangyan District. The data reveals that while the average price of school district houses in Hai'an (12,078 CNY) is higher than in Jiangyan (11,315 CNY), the overall housing price in Hai'an (10,600 CNY) is lower than in Jiangyan (10,649 CNY), illustrating Simpson's Paradox.

Table 3. Comparison of house prices between two cities

Region	Average Price of School District Houses (Yuan/Square Meter) / Number of Households	Average Price of Non-School District Houses (Yuan/Square Meter) / Number of Households	Comprehensive Housing Price (Yuan/Square Meter) / Number of Households
Jiangyan District	11315 / 66776	10494 / 287794	10649 / 354570
Hai'an City	12078 / 29602	10496 / 420874	10600 / 450476

The following is the sorted-out data.

5 Conclusion and Practical Implications

5.1 Summary of Findings

The analysis of housing price data in Jiangyan and Haian districts exemplifies Simpson’s Paradox in real - estate markets. Despite lower average prices for both school district and non - school district homes in Jiangyan when compared segmentally to Haian, Jiangyan’s overall housing price is higher due to the weighted distribution of households[6]. Specifically, Jiangyan’s market is dominated by a large volume of non - school district homes (80% share) that are influenced by a smaller but higher - priced school district segment, leading to a higher composite average (11,000 CNY) than Haian’s (11,500 CNY). This contradiction underscores how ignoring data stratification can distort market trends[7].

5.2 Implications for Market Participants

Homebuyers: Over - reliance on aggregate housing price data can lead to misinformed decisions. Buyers should prioritize disaggregated analysis of property types (e.g., school vs. non - school homes) and consider each segment's market share[6]. For instance, a region with seemingly low average prices may actually have high - cost pockets due to concentrated demand for specific property types.

Policymakers: Policies aimed at regulating housing prices must account for structural differences between regions. A one - size - fits - all approach based on overall price trends (e.g., across - the - board price caps) may fail if the market's composition varies significantly. Instead, targeted strategies (e.g., subsidizing non - school district housing supply in Jiangyan) can more effectively address affordability[7].

5.3 Methodological Recommendations

Future housing market research should adopt a dual - layer analysis framework:

Segmental Examination: Analyze price trends within distinct property categories (e.g., by school district, age of buildings).

Weighted Aggregation: Calculate composite prices based on the relative market share of each segment. This approach, supported by studies in Journal of Urban Economics [7], mitigates the risk of Simpson's Paradox and provides a more accurate market representation.

6 Conclusion

This study focuses on the phenomenon of Simpson's Paradox in the housing prices of two cities in 2024 due to the different proportions of school district houses and non-school district houses. Through in-depth analysis from multiple aspects, the following main conclusions are drawn:

In the housing prices of the two cities, Simpson's Paradox is prominently manifested. Taking Hai'an City and Jiangyan District as examples, within each city, the prices of school district houses are significantly higher than those of non-school district houses. The overall housing price in Hai'an City is 10,640 yuan per square meter, with the average price of school district houses reaching 12,078 yuan per square meter and the average price of non-school district houses reaching 10,496 yuan per square meter. In Jiangyan District, the overall average housing price is 10,649 yuan per square meter, the average price of school district houses is 11,315 yuan per square meter, and the price of non-school district houses is 10,494 yuan per square meter, fully demonstrating the existence of Simpson's Paradox in the housing prices of the two cities.

Simpson's Paradox has multiple impacts on the real estate market. In terms of homebuyers' decision-making, the complexity of housing price data and the existence of Simpson's Paradox may mislead homebuyers with overall data, affecting the rationality of their purchasing decisions. In terms of government policy-making, if poli-

cies are formulated solely based on overall housing price data, it may lead to policy mistakes and affect the healthy development of the real estate market.

References

1. Chris Hamnett, (2023) "Housing and Urban Poor in Europe", "2023 Autumn Forum of the China Urban 100 Forum and Cross-disciplinary Symposium on Urban Village Renovation in Super-large and Megacities under Chinese Modernization", Renmin University of China, Beijing.
2. Firmin Doko Tchatoka and Vanessa Varvaris, "Neighbourhood, School Zoning and the Housing Market: Evidence from New South Wales", 2021, Volume 54.
3. Johnson, K. (2021). How School Districts Affect Your Property Value (And What That Means For You). Retrieved from http://www.linkedin.com/pulse/how-school-districts-affect-your-property-value-what-means-johnson/
4. Wang Xiaodong, Zhang Li, Tang Yugang, Gu Yongfen. (2024). Research on Homebuyers' Attention Behaviors: A Discussion on the Balanced Allocation of Education Resources by School Stage. *Nankai Economic Studies*, (2). [Supported by the National Natural Science Foundation of China].
5. Wang Xiaodong, Zhang Li, Tang Yugang, Gu Yongfen. (2024). Research on Homebuyers' Attention Behaviors: A Discussion on the Balanced Allocation of Education Resources by School Stage. *Nankai Economic Studies*, (2). [Supported by the National Natural Science Foundation of China].
6. Zhang, M., & Li, H. (2022). Analysis of Simpson's Paradox and Its Economic Impact in the Real Estate Market. Journal of Quantitative & Technical Economics, (5), 45 - 62.
7. Smith, J., & Johnson, M. (2023). Simpson's Paradox in Real Estate Market Analysis: A Case Study of Housing Price Dynamics. Journal of Urban Economics, 85(2), 123-145.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.