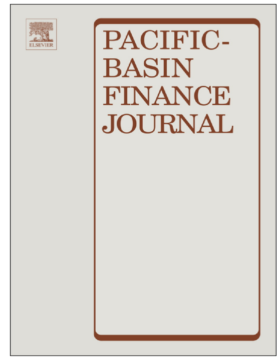
Journal Pre-proof



Exploring the value of green: The impact factors on China's second-hand green housing prices based on geographically weighted Lasso regressions

Qianwen Li, Tingyu Qian, Hui Wang, Chuanwang Sun

PII: S0927-538X(24)00413-X

DOI: https://doi.org/10.1016/j.pacfin.2024.102661

Reference: PACFIN 102661

To appear in: *Pacific-Basin Finance Journal*

Received date: 28 October 2024

Revised date: 9 November 2024

Accepted date: 29 December 2024

Exploring the value of green: The impact factors on China's second-hand green housing prices based on geographically weighted Lasso regressions

Qianwen Li[a,b], Tingyu Qian[a], Hui Wang[c], Chuanwang Sun[d,e,*]
[a] School of Business, Jiangnan University, Wuxi, Jiangsu Province, 214122, China
[b] The Institute for National Security and Green Development, Jiangnan University, Wuxi, Jiangsu Province, 214122, China
[c] Department of Econometrics and Business Statistics, Monash Busines School, Monash University, 20 Chancellors Walk, Wellington Rd, Clayton, VIC 3800, Australia
[d] China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, Fujian 361005, China
[e] MOE Key Laboratory of Econometrics, School of Economics, Xiamen University, Xiamen, Fuian 361005, China

* Correspondence to:
C. Sun, China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.
E-mail addresses: cw_sun@foxmail.com(C. Sun).

E-Mail and institutional websites：
Qianwen Li：8202108001@jiangnan.edu.cn
Institutional profile page: https://busi.jiangnan.edu.cn/info/1046/11824.htm
ORCID：0000-0001-5117-2695

Tingyu Qian：6220903008@stu.jiangnan.edu.cn
（There is no institutional websites for postgraduate students.）
ORCID：0009-0004-2691-1926

Hui Wang：hui.wang1@monash.edu
（There is no institutional websites for postgraduate students.）
ORCID：0009-0006-0095-0243

Chuanwang Sun：scw@xmu.edu.cn
Institutional profile page: http://faculty.xmu.edu.cn/SCW/zh_CN/index.htm
ORCID：0000-0002-1900-4699

Abstract:
Green housing development has progressed over the past two decades; however, the pricing advantages and influencing factors remain inadequately defined. Particularly, the effects of green housing ratings on prices and purchasing decisions have not been thoroughly researched. A dataset comprising 14,335 items (4,101 groups) was compiled of second-hand green housing transactions from spatial and temporal dimensions across various Chinese cities. To manage data clustering and enhance model robustness, the Bootstrap algorithm sampling and geographically weighted Lasso regression were utilized. The findings reveal several insights: (1) The spatial dimension notably impacts second-hand green housing prices, with regional differences evident in the effects of identical variables. This suggests that policy should be locally adapted, requiring nuanced and differentiated regulatory strategies. (2) Macroeconomic indicators, such as Gross Domestic Product, Per Capita Disposable Income, and residential commercial property sales, positively influence housing prices. Monitoring these economic indicators for timely policy adjustments is advised. (3) At the microeconomic level, the architectural features of second-hand green housing negatively affect prices in the northeast and southwest regions. Conversely, neighborhood characteristics negatively impact prices in the southeast coastal region but positively influence them in central and northeastern regions. These results suggest that regular assessments of neighborhood characteristics and stringent regulation of architectural features by the government are necessary to maintain housing stock quality. This research offers enhanced insights into price formation in the second-hand green housing market and presents vital evidence for precise policy formulation and sustainable real estate development.

Keywords:
second-hand green housing, geographically weighted Lasso regression, spatial variation, green star rating

1.Introduction

Green housing represents an innovative housing model certified by the China Green Building Evaluation Standard. This model facilitates the efficient circulation of material and energy systems, both inside and outside the house, resulting in minimal waste and pollution while achieving a certain degree of energy self-sufficiency. As the largest energy consumer (Shen et al., 2023) and carbon emitter (Xu, 2023), China has the largest construction market in the world, making energy saving and carbon reduction particularly important. By 2023, over 3 billion square meters of green building projects in China will have received national standard green building certification. This will gradually establish a green building industry characterized by energy saving, environmental protection, and high efficiency, enabling China's construction industry to achieve a higher level of greenness. Years of sustained development and policy promotion have led many property companies (e.g., China Overseas Land, Vanke Group, Greenland Group) to enter the field of green housing. This has promoted the maturity and application of green housing technology and increased the penetration of green housing in the second-hand housing market.

With the deep adjustment of China's property market, housing has entered the era of stock. The supply of new housing in the core of most cities is becoming increasingly scarce, necessitating reliance on a more resilient stock of housing to fulfill demand (source: China's New Residential Development Report 2022). The market's focus is gradually shifting from new housing to second-hand housing (Sterbenk et al., 2022; Zhang et al., 2023a). Therefore, second-hand housing is becoming increasingly important for housing supply. The trend towards 100% green building is evident, as major property companies have declared 8,145 green building labels over the past 15 years. These comprise 68% one-star, 29% two-star, and 3% three-star ratings (China Green Real Estate Development Report, 2020).This indicates increasing market recognition of green housing. As this trend develops, the share of green housing in the second-hand property market is gradually increasing. This study defines second-hand green housing as green housing that is not sold for the first time, including second-hand and multiple resale housing certified to green standards.

As China's property market increasingly enters the era of inventory, homebuyers' concern for house prices remains central to their decision-making, whether for first-time sales or multiple resale houses (Hsu, 2013). The dominant factor affecting the first sale price of green housing remains the star rating. There is a significant positive correlation between the star rating of green housing and the transaction price (Dell and Bottero, 2021; Li et al., 2021a). Specifically, compared to non-green housing, green housing with national standard green building certification is generally sold at higher prices. The price premium is more pronounced as the star rating increases (Jiang et al., 2021). In the second-hand housing market, it is uncertain whether the green star rating remains a key variable influencing house purchase decisions and house prices, and whether it reflects the market's recognition of the sustainable performance of green housing.

Current research has under-explored the factors influencing the price of second-hand green housing, particularly the impact of green star ratings. Star ratings for green housing are a key factor in measuring the performance of green housing and influencing its promotion. This gap limits the effective promotion of green sustainable development in the second-hand housing market. Additionally, second-hand green housing prices are closely related to factors such as architectural features and neighborhood characteristics. Since price is a core factor constraining the universal promotion of high-value green products, it is crucial to investigate the core factors influencing the second-hand prices of green housing. This study systematically extends the theoretical framework for analyzing the second-hand prices of green housing based on existing literature. It comprehensively analyzes the multiple factors affecting house prices, ranging from macro real estate market sentiment and location economy to micro building characteristics and neighborhood environment. This approach not only enriches the theory of real estate economics but also provides market participants with a more scientific basis for decision-making, which is significant in promoting the green and sustainable development of the real estate market.

The innovation of this study includes: (1) At the theoretical level, it considers the impact of real estate market prosperity, locational factors, and economic factors on second-hand green housing prices from a macro perspective. It also accurately measures the role of foundational elements such as architectural features and neighborhood characteristics from a micro perspective. It further explores the regional variability of different factors, expanding the price econometric framework of the second-hand green housing market. (2) At the methodological level, this study conducts empirical analysis of nearly 20 years of transaction data from China's second-hand green housing market, gathered via big data multi-channel scraping. It employs the Bootstrap algorithm for repeated sampling to avoid data clustering issues and introduces Lasso regression into the GWR model. While considering regional differences, it effectively addresses collinearity issues, enhancing the model's predictive accuracy and stability. This provides a new methodological framework for analyzing factors affecting prices in the second-hand green housing market.

The subsequent chapters are arranged as follows: The second section provides a literature review of the main factors affecting second-hand green housing prices and the geographic regression model (GWR) and its expansion. The third and fourth sections cover data collection and research design, including large-scale crawling of relevant data, repeated cleaning, use of various models, and selective empirical analysis to explore factors influencing green housing prices in different regions. The fifth section compares the results, and the sixth section presents conclusions and policy insights for the sustainable development of the green housing second-hand market.

2.Literature review

2.1 House price influencing factors

In the analysis of factors influencing housing prices, substantial differences exist between Western and Chinese countries. These differences are evident not only in economic development levels, cultural backgrounds, and social systems but also profoundly affect the operational mechanisms and price formation processes of their respective housing markets. The Chinese government regulates market supply and demand through frequent and direct policy interventions, such as restrictions on purchases, loans, and land supply (Jia et al., 2018; Chen et al., 2019). This approach contrasts with the Western reliance on market self-regulation and free competition. In terms of consumer behavior, Chinese residents display distinct home-buying preferences, such as a strong interest in school districts (Huang et al., 2020; Peng et al., 2021), whereas Western residents typically prioritize accessibility and quality of life. Additionally, factors indirectly influencing housing price volatility are significant. For instance, real estate market bubbles in China are directly linked to housing price trends, whereas in the West, they are more commonly tied to speculative behavior and market expectations (Zhang et al., 2021). The willingness to pay and investment behavior patterns of residents (Hudson, 2024) critically affect pricing in the Chinese second-hand housing market. Moreover, the strategies and adjustments of financial institutions (Nguyen et al., 2022; Bedendo et al., 2023) significantly influence the second-hand housing market through mechanisms including market supply and demand regulation, impact on market expectations, and capital flow alterations.

As early as 2005, Kleindorfer and other scholars suggested the importance of green design for sustainable development and that green labels derived from green design will directly affect market prices (Agatz et al., 2021). Islam et al. (2022) systematically analyzed the impacts of star rating on the environmental impacts and costs of a green housing building over its entire life cycle (construction, operation, maintenance, and disposal). They assessed environmental and economic well-being in an integrated manner, suggesting that star ratings are not only about cost but also a key basis for selecting environmentally friendly materials and construction methods and optimizing the environmental performance of a building's lifecycle. Naderi et al. (2022) found that star ratings significantly impact the effectiveness of residential pre-cooling in Australia. High-star-rated buildings provide more effective cost savings, peak demand reduction, and thermal comfort enhancement than lower-star ratings and are particularly valuable in reducing grid loads. Safarova et al. (2022) conclude that building star ratings significantly affect the energy efficiency and occupant comfort of housing and emphasize the importance of star ratings in improving building performance. Naderi et al. (2023) used clustering analysis to obtain significant energy cost savings for residential summers with star ratings.

In the real estate market, the star premium of green housing directly affects its transaction price. Porumb et al. (2020) confirmed that the price premium of green housing is 19% higher than ordinary housing and shows regional differences through the study of the costs and advantages of green certification. Dell and Bottero (2021) used a hedonic price model to confirm a green premium for green buildings in Singapore. Jiang et al. (2021) used the same model to confirm that the price premium of certified green housing in China is 6% at the resale stage. The higher the star rating, the higher the premium. Li et al. (2021b) confirmed that the rental premium of LEED-certified housing is 19.5%, with regional differences in the price premium. Yuan et al. (2022) used a hedonic pricing model to calculate that the average discount on green housing in the second-hand market is 11.94%. The higher the house price, the larger the discount. Hui et al. (2021) compared the willingness to pay for green housing with luxury housing in Hong Kong. They found that consumers are more willing to pay a higher premium for certified green housing.

The Gross Domestic Product (Bednarek et al., 2021; Yii et al., 2022), residents' income (van et al., 2023), consumption, regional differences (Zheng et al., 2023), and property market conditions significantly impact second-hand housing prices. André et al. (2021) collected data from 1,338 residents in China and found that the willingness to purchase green housing is positively related to household assets. Chen et al. (2023) used a machine learning method to compare housing conditions under different conditions and found that regional variation in economic factors was more significant. This suggests that price dynamics in the property market are closely linked to macroeconomic indicators, reflecting the combination of economic, social, and technological factors that shape the market's direction.

Huang (2023) used data from green housing transactions in Taiwan to explore the impact of green buildings on house prices. The results showed that the number of rooms, floor height, and age of the housing significantly impact house prices, with green certification having a relatively greater impact. Porumb et al. (2020) used a spatial hedonic model to measure that closer proximity to transportation increases house prices. Qu et al. (2020) analyzed residential house prices in Wuhan, China, revealing the positive impact of hospitals, parks, and universities on house prices. They also found that traffic and plot ratio negatively impact house prices, with different impacts in different regions. Yang et al. (2020) explored the impacts of the BRT system on house prices in Xiamen, China, suggesting that transport accessibility causes housing premiums but also a proximity penalty. Han et al. (2021) used Shenzhen residence prices as the dependent variable and confirmed the significance of culture, green areas, shopping malls, and central business districts. Wang and Li (2022) found that residential premiums are higher in high-quality school districts, illustrating the positive effect of educational proximity on house prices. Chen et al. (2022) argue that medical accessibility is a key factor affecting house prices. This suggests that a building's characteristics (floors, number of rooms, orientation, age of housing (Jiro,2020)) and neighborhood characteristics (education, transport, living, healthcare) significantly impact house prices.

2.2 GWR model

Brunsdon et al. (1996) proposed the geographically weighted regression (GWR) technique, which integrates Tobler's (1970) first law of geography into local spatial statistical methods. Spatial regression coefficients correspond one-to-one to spatial location, and parameter estimates vary with spatial location. These estimates are obtained by performing regression analysis modeling for each independently sampled analysis point. GWR is a local spatial regression method used in various fields such as meteorology, environmental management, and ecology (Tasyurek and Celik, 2022). To further explore the application of GWR in real estate research, scholars have conducted a series of empirical studies. Liu et al. (2019) used OLS and GWR to explore the factors influencing land ecological security and their spatial relationships. Wu (2020) analyzed the evolution of the ecological footprint in China's provinces using time-series data. It was found that the GWR model had advantages in regression goodness-of-fit, variance comparison, and residual spatial autocorrelation compared to the OLS model.

Location is a crucial factor in house prices. Proximity to city centers, quality of neighborhoods, and proximity to parks and attractions significantly affect house prices. The GWR model can capture the micro-mechanisms of house price formation more accurately. It adds location information of house price observations at different latitudes and longitudes to the regression model and uses spatial autocorrelation to explain changes in the observations. Li et al. (2016) used GWR to reveal spatial heterogeneity in the effects of forest cover and air pollution on house prices. Wang et al. (2021) used the GWR model to quantitatively analyze the impact of road traffic on urban air quality. Jin et al. (2023) similarly used the GWR model and found a significant relationship between house prices and the green landscape index with spatial effects. Therefore, the GWR model, by incorporating geospatial characteristics and reproducing geographic factors in the process of house price formation, has a wide range of applications in house price analysis and assists in explaining the formation mechanism of house prices.

The GWR model is crucial in spatial data analysis, especially for data with spatial dependence and heterogeneity. However, the problem of collinearity limits its application. Several scholars have combined other methods to overcome this challenge in recent years. Chen et al. (2020) proposed the geographically weighted ridge regression (GWRR) method, which addresses the collinearity problem and improves prediction accuracy by introducing ridge regression to enhance the GWR model. Xu et al. (2021) proposed a multifactor geographically weighted machine learning (MFGWML) algorithm for the dimensionality reduction of surface temperature (LST) data. This algorithm combines XGBoost, multivariate adaptive regression spline (MARS), and Bayesian Ridge Regression (BRR). The results show that MFGWML outperforms other classical algorithms in the dimensionality reduction of LST data. He et al. (2021) proposed the adapted geographically weighted Lasso (Ada-GWL) framework. Through empirical analyses of the Shenzhen Metro, the Ada-GWL model performs best in terms of estimation error and goodness-of-fit. Murakami et al. (2023) used the Poisson regression model for spatial data and proposed linearised geographically weighted Poisson regression (L-GWPR). Monte Carlo experiments validated the accuracy and high efficiency of ridge regression regularised L-GWPR. In conclusion, GWR models combined with other methods provide richer and more accurate tools for spatial data analysis. For different application scenarios, choosing the appropriate GWR model according to data characteristics and problem demands is especially important.

3.Research design and data

As an authoritative website hosted by the Science and Technology and Industrialization Development Center of the Ministry of Housing and Urban-Rural Development of China, GreenBuild.com provides comprehensive and objective data. To study the development of green housing more deeply, the building type "housing" was screened from this website. This included one-star, two-star, and three-star green housing plots and buildings. Octopus Collector (a Chinese data collection application, version 8) was used to crawl the transacted second-hand housing data from Chain Store's official website in each city of each province in China (Li et al., 2023). This data covered 27 key indexes, such as transaction period, area, price, year, floor, number, traffic, and education distance. The 176,877 data points were screened and cleaned several times to remove datasets with unclear land parcel information and excessive missing data. The temporal and spatial distribution of the data is shown. Eventually, the dataset was grouped into two categories: architectural features and neighborhood characteristics, providing a basic dataset for subsequent in-depth analysis and modeling.

Previous studies have confirmed that macroeconomic factors, such as Gross Domestic Product (GDP), Consumer Price Index (CPI), and Per Capita Disposable Income (PCDI) (see the Literature Review section for details), affect second-hand housing prices. These factors significantly impact house prices in the current period and have a far-reaching effect on future house price trends. This study specifically included macroeconomic variables with lags in its analyses to explore their relationship with second-hand green housing prices more comprehensively. In addition to macroeconomic factors, property market sentiment is also an important factor affecting second-hand housing transactions. High or low market sentiment is directly related to trading activity and the price level of second-hand housing. This study adds the key indicator of residential commercial property sales (RCP) in each province in the current year to more accurately measure the impact of market dynamics on second-hand housing prices. The specific parameter settings are shown in Table 1.

Table 1 Parameter setting table

Nature of the variable	Variable meaning	Calculation method
Explanatory variable	Second-hand house prices	Transaction price of second-hand housing/gross floor area of second-hand housing (CNY/square meter)
Architectural features	Star	According to the Green Building Evaluation Standard (DBJ/T45-104-2020), it is divided into 1, 2, and 3 stars (the basic level is not in the scope of the study)
	Age of housing	Year sold - year built +1
	Degree of decoration	Categorical variables (hardcover = 4, shortcover = 3, rough = 2, other = 1)
	With or without a lift	Categorical variables (yes=1, no=0)
	House orientation	House orientation (due north or south = 1, other = 0)
	Story	Categorical variables (high=3, medium=2, low=1)
	Number of rooms	Total number of rooms in the rate (including toilets)
Neighborhood characteristics	Education	Average of the three closest distances in the neighborhood to kindergartens, primary schools, and secondary schools
	Transportation	Distance of the neighborhood from the nearest metro or bus
	Healthcare	Distance to the nearest hospital in the neighborhood

	Lifestyle	Total number of entertainment, shopping, banks, etc. within a 15-minute lifestyle circle (Zhang et al.,2022)of the neighborhood
Macro factors	Gross Domestic Product (GDP)	Gross Domestic Product (GDP) of each province in the year of sale (trillions of yuan)
	GDP lagged by 1 period	GDP of each province in the year before the year of sale (trillion yuan)
	GDP lagged by 2 period	GDP of each province for the two years preceding the year in which it was sold (trillions of yuan)
	GDP lagged by 3 period	GDP of each province for the three years preceding the year in which it was sold (trillions of yuan)
	Consumer Price Index (CPI)	Consumer price index for the year of sale (CNY)
	Per capita disposable income (PCDI) of the population as a whole	Per capita disposable income of all residents in each province in the year preceding the year of sale (CNY)
	Residential Commercial Property Sales (RCP) for the year	Residential commercial property sales by the province in the year sold (billion yuan)

Due to objective reasons, missing macro data in 2023 necessitates adopting different fitting methods for different data sets. 2023 data are converted using the first half of the year because residential commercial property sales in each province are not published half-yearly. Therefore, data from each province from 2003-2022 are used as reference values for the prediction, and the residential commercial property sales for 2023 are forecasted using the ARIMA model. The processed data totaled 14,335 items (4,101 groups), with each entry representing the complete data of a second-hand green housing transaction, including factors such as age, decoration, lift availability, and room count. During data selection and calculation, academic norms and the reliability of data sources were strictly followed. All data were sourced from the National Bureau of Statistics to ensure the authority and accuracy of the data. After repeated cleaning, data requiring mutation levels were forward normalized and reverse normalized to ensure consistency (i.e., the impact on house prices is numerically revised so that a larger variable corresponds to a higher house price). To provide more clarity on the actual situation of each variable, descriptive statistics are tabulated below:

Table 2 Parameter descriptive statistics table

	Mean	Maximum	Minimum	Median	Variance
Y-housing price	28257.83	123221.69	2058.38	23321.10	280633884.06
Star	2.42	3.00	1.00	3.00	0.64
RCP	6956.07	19829.63	421.85	5173.60	29050895.55
GDP	98.33	121.02	41.21	101.36	299.47
PCDI	38712.82	79610.00	15749.00	32914.00	255071897.44
CPI	28930.30	39344.00	12668.00	31013.00	28438714.23
Age	8.25	20.00	1.00	8.00	9.18

Decoration	3.22	4.00	1.00	4.00	1.47
Lift	0.98	1.00	0.00	1.00	0.02
Orientation	0.83	1.00	0.00	1.00	0.14
Story	2.05	3.00	1.00	2.00	0.63
Room	6.66	18.00	1.00	6.00	2.38
Education	575.27	1526.67	243.00	627.00	42032.79
Transportation	380.54	1469.00	115.00	318.00	51875.78
Healthcare	992.08	2500.00	206.00	1086.00	157026.94
Lifestyle	46.82	77.00	9.00	46.00	242.78

Note: PCDI denotes disposable income per capita for all residents, RCP denotes residential commercial property sales for the year, GDP denotes Gross Domestic Product, and CPI denotes Consumer Price Index.

Specific data indicators of all the variables and their calculations are shown in detail in Figure 1.

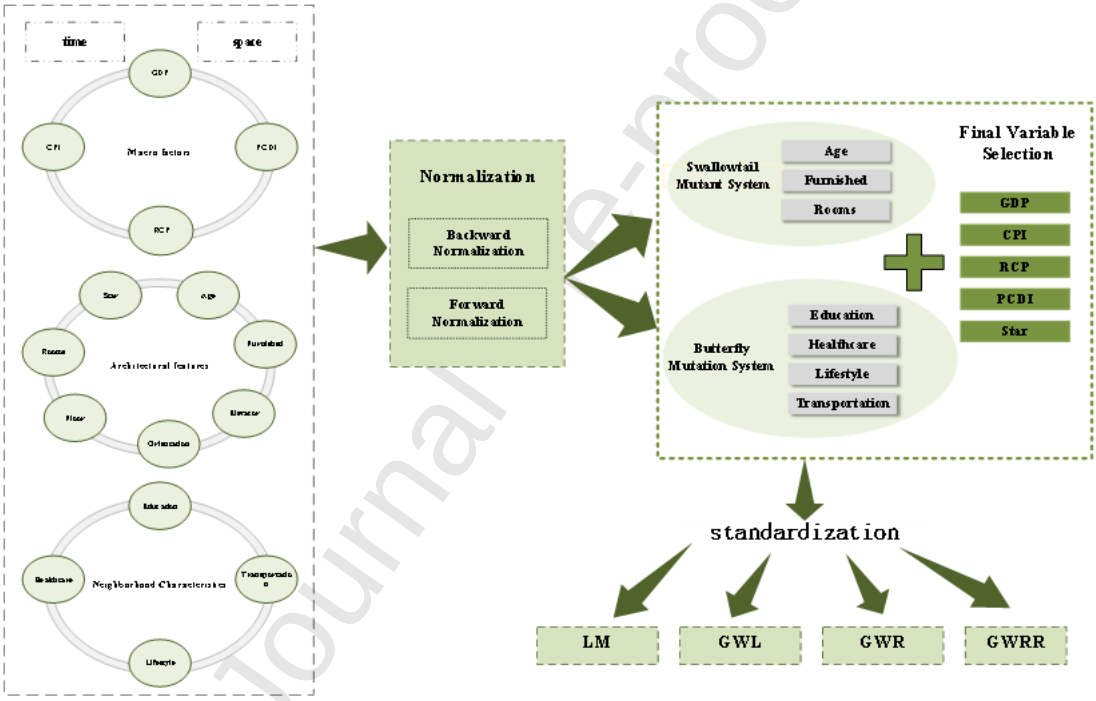


Fig.1 Variable handling process

Note: PCDI denotes disposable income per capita for all residents, RCP denotes residential commercial property sales for the year, LM denotes general linear regression, GWL denotes geographically weighted Lasso regression, GWR denotes geographically weighted regression, and GWRR denotes geographically weighted ridge regression

The normalization formula for each variable X is

$$X_n = (X - X_{min})/(X_{max} - X_{min}) \quad (1)$$

Where, X_{max} denotes the maximum value in the variable, and X_{min} denotes the minimum value in the variable X. After the classification of variables, to ensure the consistency and comparability of the data, the original data are used to standardize the data through Z-score (Urolagin et al., 2021) to unify the scale, and the formula for standardization is

$$X = (X - \mu)/(\sigma) \quad (2)$$

Where, μ is the mean value of the variable, σ is the standard deviation of the variable.

4.Model

Multiple cleaning processes resulted in second-hand green housing data (with star ratings) for 71 neighborhoods (4101 houses). Applying the traditional General Linear Regression (GWR) model neglects geospatial interactions. The GWR model has disadvantages, including data point over-aggregation, difficulty in bandwidth estimation, and slow running speed. Therefore, the Bootstrap algorithm was introduced to estimate bandwidth by randomly sampling points, conducting 1,000 iterative experiments to cover the full data range and improve model accuracy. Additionally, significant collinearity issues among some variables necessitated the use of Geographically Weighted Ridge Regression-GWRR (Pourmohammadi et al.,2021) and Geographically Weighted Lasso Regression-GWL (He et al.,2021) for effective treatment, based on Bootstrap algorithm sampling. These methods were compared with General Linear Regression-LM (Aslett et al.,2024), GWR, and errors. The model with the highest goodness-of-fit and the smallest error was selected for calculations, as shown in Figure 2.

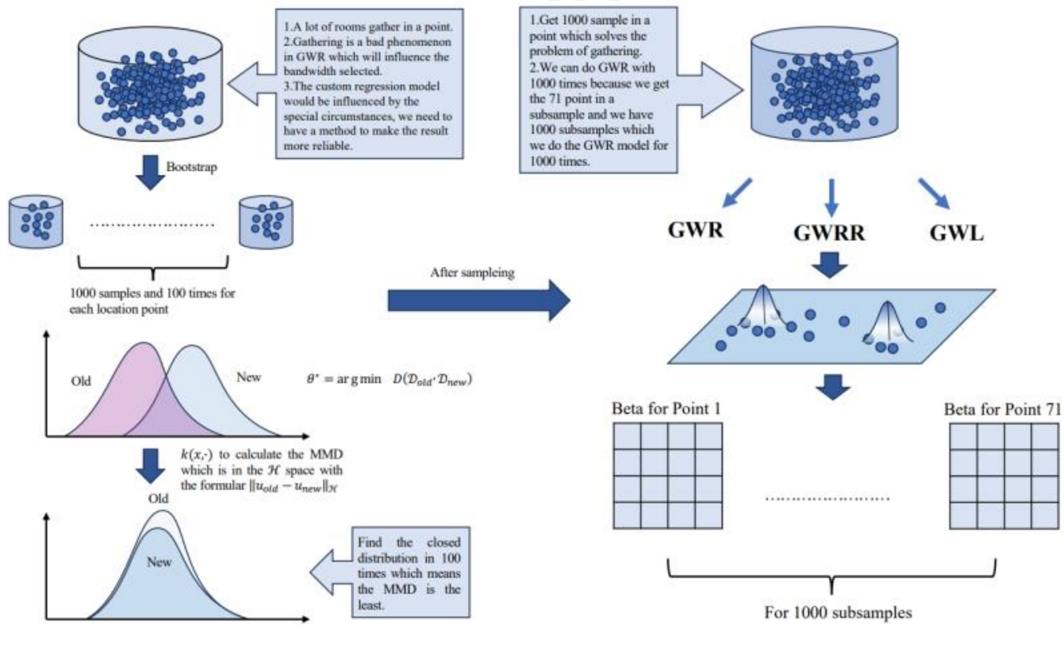


Fig.2 Model selection diagram

Note: GWL denotes geographically weighted Lasso regression, GWR denotes geographically weighted regression, and GWRR denotes geographically weighted ridge regression

The MMD in Figure 2 refers to the Maximum Mean Discrepancy, which needs to be calculated firstly as the kernel function and then the center of mass, and finally, the distance is measured by the probability integral metric, and the specific formulae are as follows：

$$\| u_{old} - u_{new} \| = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} A^T x_i - \frac{1}{n_t} \sum_{j=n_s+1}^{n_s+n_t} A^T x_j \right\|^2$$

$$= tr\left(A^T [X_s \ \ X_t] \begin{bmatrix} \frac{1}{n_s^2} 11^T & \frac{-1}{n_s n_t} 11^T \\ \frac{-1}{n_s n_t} 11^T & \frac{1}{n_t^2} 11^T \end{bmatrix} \begin{bmatrix} X_s \\ X_t \end{bmatrix} A \right) = tr(A^T X M X^T A)$$

(3)

u_{old} represents the value of the variable without the Bootstrap algorithm, while u_{new} represents the corresponding value in the model with the Bootstrap algorithm. If used in kernel trick ($\phi(x)$), then

$$\left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_j) \right\|^2 = tr\left([\phi(x_s) \ \ \phi(x_t)] \begin{bmatrix} \frac{1}{n_s^2} 11^T & \frac{-1}{n_s n_t} 11^T \\ \frac{-1}{n_s n_t} 11^T & \frac{1}{n_t^2} 11^T \end{bmatrix} \begin{bmatrix} \phi(x_s)^T \\ \phi(x_t)^T \end{bmatrix} \right)$$

$$= tr\left(\begin{bmatrix} < \phi(x_s), \phi(x_s) > & < \phi(x_s), \phi(x_t) > \\ < \phi(x_t), \phi(x_s) > & < \phi(x_t), \phi(x_t) > \end{bmatrix} M \right) = tr\left(\begin{bmatrix} K_{s,s} & K_{s,t} \\ K_{t,s} & K_{t,t} \end{bmatrix} M \right) \quad (4)$$

Where

$$(M)_{ij} = \begin{cases} \frac{1}{n_s n_s}, & x_i, x_j \in \mathcal{D}_s \\ \frac{1}{n_t n_t}, & x_i, x_j \in \mathcal{D}_t \\ \frac{-1}{n_s n_t}, & \text{otherwise} \end{cases} \quad (5)$$

Where x denotes the source domain (mapped to the target domain by the kernel function), n denotes the data length, and A denotes the kernel function, $\mathcal{D}_s$ denotes the data distribution of the original data, $\mathcal{D}_t$ denotes the data distribution after Bootstrap sampling.

The GWR model in this study is formulated in OLS (least squares estimation) as follows:

$$\ln(y_i) = \beta_0(u_i, v_i) + \sum_{j=1}^{n} \beta_j (u_i, v_i) x_{ij} + \varepsilon_i \qquad i = 1.2\ldots4101, \ n = 7$$

(6)

Where i indicates 4101 group data labels, y_i indicates unit rates for each group (CNY/m^2), β_j indicates estimated coefficients for each variable, ε_i denotes the random error for group i, (u_i, v_i) denotes the latitude and longitude of data set i, j denotes the jth parameter, x_{ij} denotes the jth data in group i, including GDP, CPI, RCP, PCDA, Star, and two mutation level variables (architectural features, neighborhood characteristics). Using the R software in the Ubuntu server system on a 4-core 24G ARM architecture CPU, the model results confirm that the model fit goodness of GWL (i.e., GWR + Lasso) has a high probability distribution with minimal error.

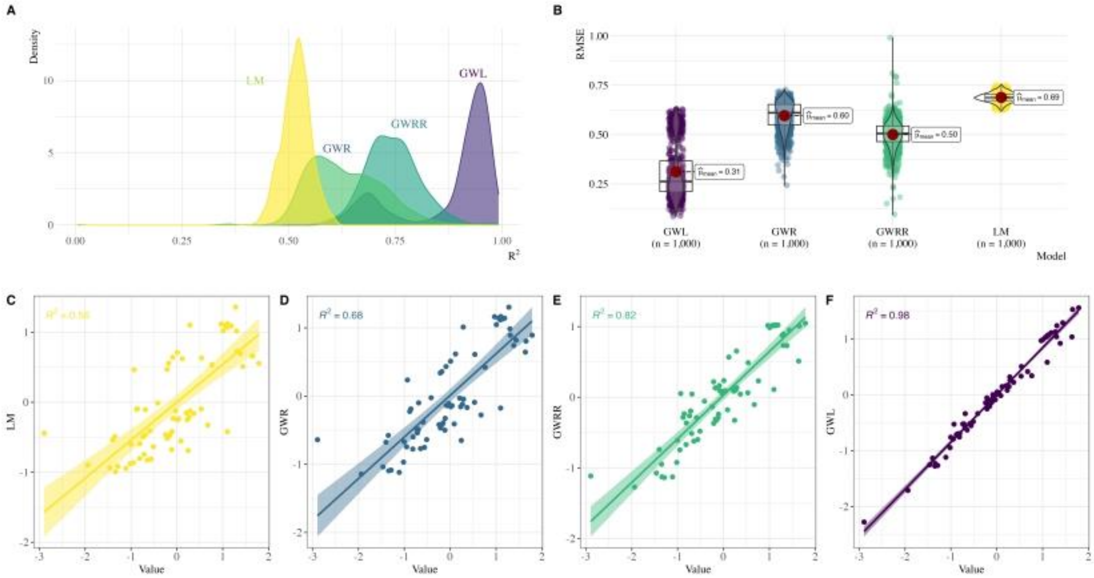


Fig.3 Comparison of the goodness of fit and error of each model

Note: LM denotes general linear regression, GWL denotes geographically weighted Lasso regression, GWR denotes geographically weighted regression, and GWRR denotes geographically weighted ridge regression.

As shown in Fig. 3a, the goodness-of-fit distribution of the 1000 data sets for GWL exceeds 80%, which is significantly higher than the other models. Fig. 3b illustrates that the mean value and distribution interval of the root-mean-square error for GWL are lower than those of the other models, indicating that GWL has the optimal fit and the smallest relative error. From Fig. 3c-f, the average goodness-of-fit across 1000 groups follows the order LM < GWR < GWRR < GWL. Additionally, the average distance of the 1000 points to the shortest diagonal is also the smallest for GWL, indicating higher relative accuracy. Therefore, GWL is selected as the final model in this study.

GWL combines Lasso regression and GWR models, incorporating an additional penalty factor in the Lasso regression model. This factor reduces values in the loss parameter through regularization. The Lasso(Lucey et al.,2022; Nazemi et al.,2018; Zhang et al.,2019) estimates are as follows:

$$\hat{\beta}_{\mathrm{Lasso}} = \arg\min_{\beta}\left[\sum_{i=1}^{m}\left((\ln(y_i) - \beta_0(u_i, v_i) - \sum_{j=1}^{n}\beta_j(u_i, v_i)x_{ij})^2\right) + \lambda\sum_{j=1}^{n}\left|\beta_j\right|\right] \quad (7)$$

At this point, when the parameters of the adjustment coefficient are less than $\sum_{j=1}^{k}\left|\widehat{\beta}_j(\mathrm{OLS})\right|$ will be compressed, which in turn reduces the complexity of the model, and in the choice of penalty coefficients, the use of traversal algorithm, to 0-1 as the interval, step 0.01 for traversal, select the root mean square error of the smallest penalty coefficients.

5.Analysis results

As shown in Fig. 3, the parameter indicators of each model type are presented in

Table 2. The GWL model exhibits the highest goodness-of-fit (R-Square) and the smallest Root-Mean-Square Error of Prediction (RMSE) and Root-Mean-Square Percentage Error (RMSPE). Among the four models, GWL aligns most closely with the data and has the lowest error rate. Additionally, GWR effectively improves the goodness-of-fit of LM, indicating that the geographic correlation between second-hand green housing prices is significant. Different models combined with GWR also show varying optimal bandwidths, with GWRR and GWL having relatively small bandwidths.

Table 3 Comparison of the parameters of each type of model

	LM	GWR	GWRR	GWL
R-Square	0.5181753	0.6339942	0.7406643	0.8812502
RMSE	0.6888321	0.5963994	0.50067	0.311995
RMSPE	0.6888321	0.771986	0.7568408	0.5300464
Bandwidth	-	16.47982	3.917158	7.169776

Note: LM denotes general linear regression, GWL denotes geographically weighted Lasso regression, GWR denotes geographically weighted regression, and GWRR denotes geographically weighted ridge regression.

The results of each parameter of the GWL model are shown in Table 3. Overall, the values of each variable at each point vary significantly, displaying both positive and negative impacts. The mean value indicates that a star rating has a positive impact on house price; other conditions being equal, a higher star rating corresponds to a higher house price. Macroeconomic factors also affect house prices, with RCP, CPI, and GDP having positive impacts, while PCDI has a negative impact. Architectural features and neighborhood characteristics generally show negative impacts. In terms of standard deviation, the national PCDI has the largest variance, and star ratings show the highest consistency (with less volatility in the data). From a median perspective, across the nation, only architectural features and the Consumer Price Index (CPI) negatively impact housing prices, while all other variables have a positive effect. From a max-min value standpoint, housing prices in certain regions are significantly influenced by provincial macroeconomic factors.

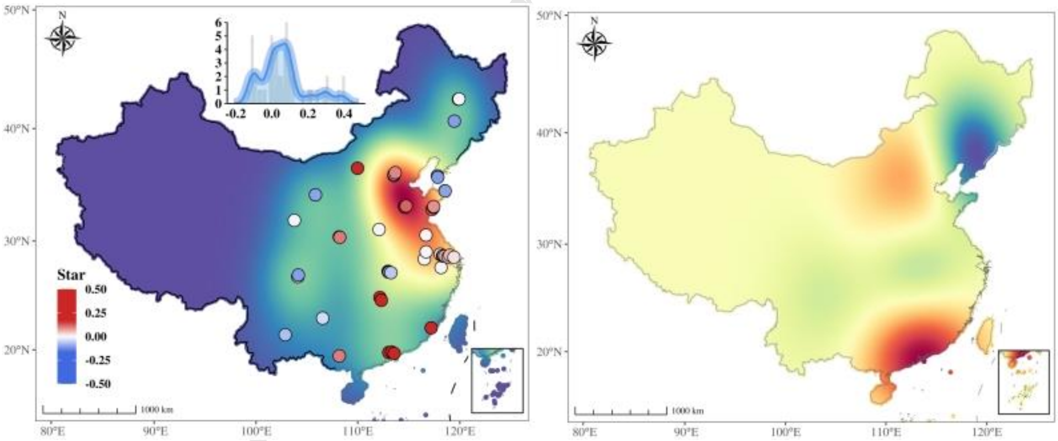
Table 4 Analysis of GWL model parameters

	MEAN	STANDARD DEVIATION	MEDIAN	MIN	MAX
β_0	0.074463	0.625604	0.008023	-2.061912	3.053336
star	0.068307	0.175270	0.047759	-0.423870	0.756836
RCP	0.008404	0.687401	0.033206	-2.597291	3.239125
PCDI	-0.143275	1.250689	0.181249	-9.091472	0.720611

CPI	0.054757	0.525008	-0.033018	-0.272523	3.832870
GDP	0.113079	0.440346	0.050326	-1.180727	1.661727
neighborhood characteristics	-0.167538	0.684938	0.000622	-3.427176	0.521016
architectural features	-0.001802	0.310356	-0.028641	-0.734881	1.910002

Note: PCDI denotes disposable income per capita for all residents, RCP denotes residential commercial property sales for the year, GDP denotes Gross Domestic Product, and CPI denotes Consumer Price Index.

From the above analysis, it is clear that regional differences cause changes in the parameters of variables, making it essential to analyze these differences. The geospatial distribution of star ratings is shown in Figure 4. In Figure 4a, colors represent the density of second-hand green housing with star ratings: purple indicates no data, and darker colors show higher density.



a.Star Rating Map b.Comparison of Interpolated Distribution of Starred Kriging

Fig.4 Comparison of the distribution of national star rating coefficients

Note: the colors in Figure 4.a represent the density of second-hand green housing with star ratings, with purple indicating the absence of data. The darker the other colors, the more densely packed the area is with second-hand housing. Figure 4.b, using Kriging interpolation for its distribution, demonstrates that the deeper the red, the more significant the positive impact of the indicator on local housing prices, while the deeper the blue, the more substantial the negative impact.

As seen in Figure 4, there is a decrease in the number of second-hand green housing with star ratings from the coastal regions (including the Bohai Economic Rim and the Yangtze River Delta Economic Zone) to inland areas. Among the 71 communities surveyed, the star rating coefficient is most densely distributed around 0.1 $\pm$0.05, indicating a predominantly positive impact on second-hand green housing prices. In the southeastern coastal regions, a higher star rating for green housing correlates with higher second-hand green housing prices. Conversely, in the central and northeastern regions, the star rating has a negative impact on housing prices, a trend associated with macroeconomic factors, as detailed in Figure 5.

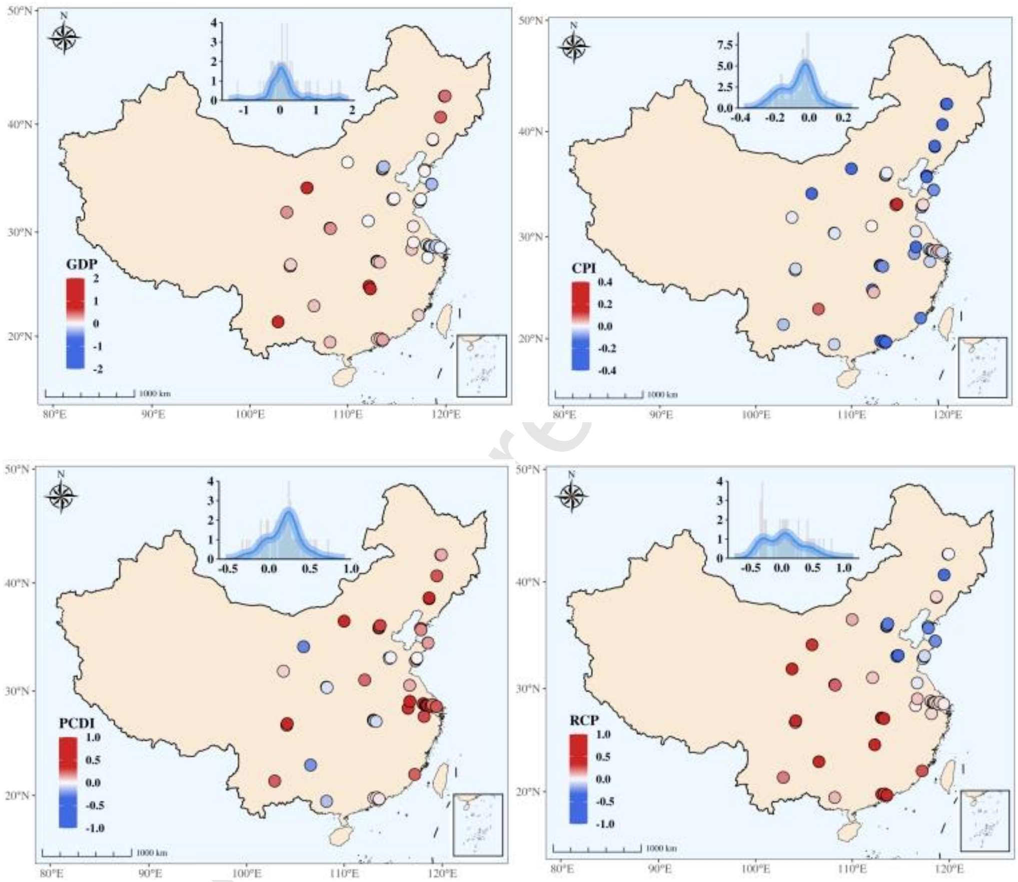


Fig.5 Comparison of the distribution of macro coefficients

Figure 5 illustrates the macroeconomic influences on housing prices. Gross Domestic Product (GDP), Per Capita Disposable Income (PCDI), and Residential Construction Permits (RCP) generally have a positive impact on housing prices across various regions, while the Consumer Price Index (CPI) tends to have a negative effect on second-hand housing prices. Regions with higher GDP rankings show a comparatively minor impact on housing prices. The CPI coefficient is predominantly negative across regions because CPI does not include housing expenditure. When total assets are limited, increased consumption expenditure leaves fewer assets available for purchasing property, thus creating a negative relationship. The PCDI coefficient is largely positive across regions, indicating that an increase in per capita disposable income, and consequently long-term assets, positively affects housing prices. The correlation between RCP and housing prices is weaker in the northern regions, while in other regions, a thriving real estate market correlates with higher second-hand housing prices.

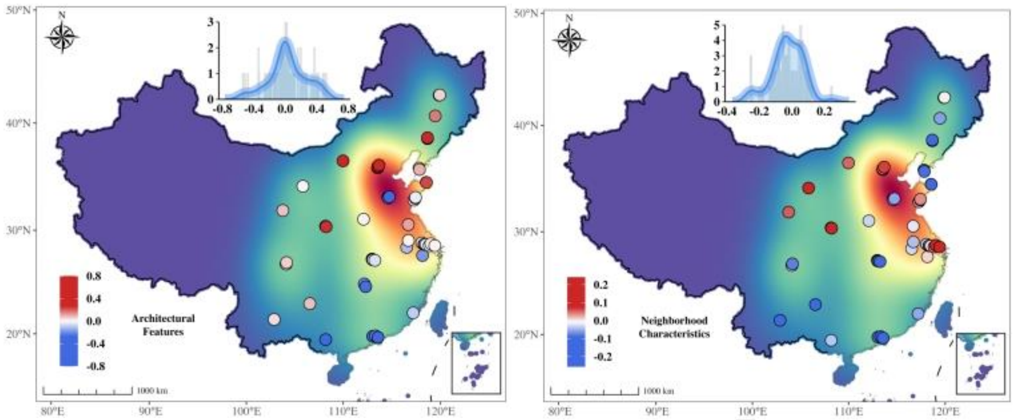


Fig.6 Comparison of the distribution of architectural features, neighborhood characteristics coefficients across the country

Note: The colors of the map represent the density of green housing second-hand properties with stars, purple represents the absence of data, and the darker the rest of the colors, the denser the concentration of second-hand housing in the area.

Regarding architectural features, most coefficients cluster around 0 ± 0.1. In China's northeastern and southwestern regions, architectural features (such as building age, decoration level, and number of rooms) negatively impact housing prices. In the central and eastern coastal regions, these features positively influence second-hand housing prices. Concerning neighborhood characteristics, most coefficients gather around 0 ± 0.2. In the southeastern coastal areas, better neighborhood characteristics (proximity to education and medical facilities, convenient transportation, and comprehensive infrastructure) are associated with lower second-hand housing prices. Conversely, in the central and northeastern regions, better neighborhood characteristics correlate with higher housing prices.

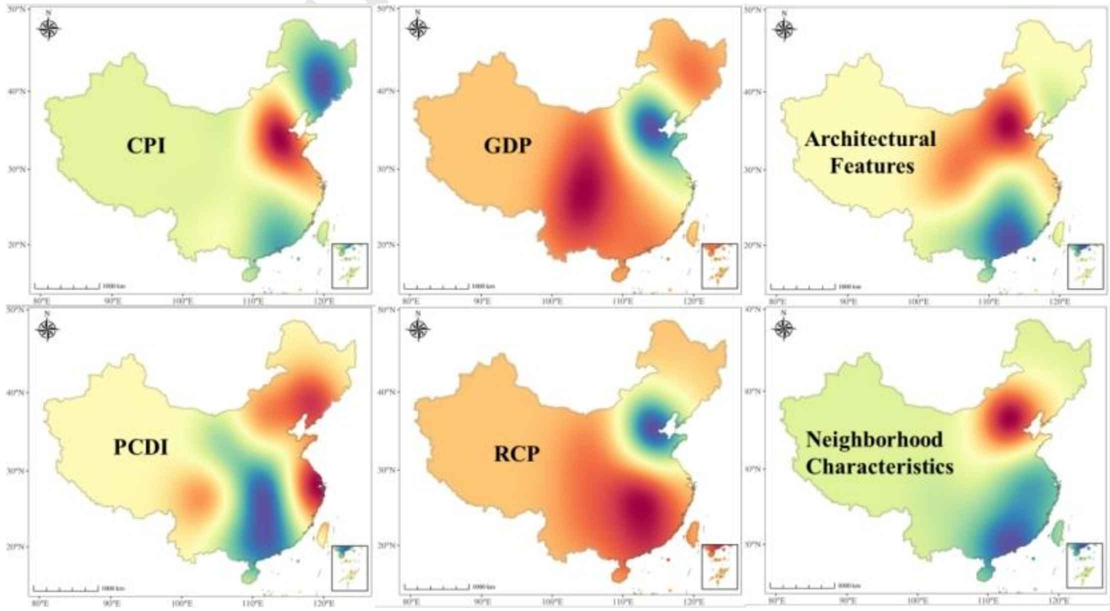


Fig.7 Comparison of the spatial distribution of different coefficients with the Kriging interpolated

distribution

Note: The darker the red color, the more significant the positive impact of the indicator on house prices in the region, and the darker the blue color, the more significant the negative impact.

The color distribution of the six parameter coefficients indicates considerable spatial variation in their impact, yet there is also consistency across parameters. The spatial impact of GDP and RCP on housing prices is broadly consistent. In China's Bohai Economic Rim, GDP and RCP negatively impact housing prices, whereas in other regions, they have a positive effect. Architectural features and neighborhood characteristics show a similar pattern. In the southeastern coastal regions of China, these factors negatively influence housing prices, while in the Bohai Economic Rim, they positively affect prices. In Beijing, Tianjin, Hebei, Shandong, and Jiangsu, the CPI positively impacts housing prices. However, in the northeastern regions and along the southeastern coast, the CPI negatively affects housing prices. In the Bohai Economic Rim and the Yangtze River Delta Economic Zone, PCDI positively influences housing prices, whereas in central China, Guangdong, and Guangxi, PCDI has a negative impact.

To prevent the heterogeneity of the results, this study used the GWR and GWRR for robustness testing to ensure the stability of the calculations. Since the LM model is not spatial and the resulting coefficients are only one set, they are not compared. The results are presented in the table below.

Table 5 Table of coefficients for multi-model robustness tests (comparison of means)

Model	star	RCP	PCDI	CPI	GDP	neighborhood characteristics	architectural features
GWL	0.0683	0.0084	-0.1433	0.0548	0.1131	-0.1675	-0.0018
GWR	-0.0481	0.0948	0.6534	-0.2048	0.1268	0.0406	-0.0493
GWRR	0.0153	0.0635	0.3091	-0.0830	0.0350	-0.0016	-0.0147

Note: GWL denotes geographically weighted Lasso regression, GWR denotes geographically weighted regression, and GWRR denotes geographically weighted ridge regression.

From the table, it can be seen that the results calculated by GWL are robust, and the differences in parameter estimates are not significant. At least two of the coefficients are consistent across the three groups of models, except for CPI. Compared to the GWRR, there is 88.7 percent positive and negative agreement for all values of this coefficient at the star. Therefore, the final calculations of the GWL model are robust.

6.Conclusion

6.1 Discussion and contribution

This study examines the prices of second-hand green housing by extracting transaction data for 4,101 units across various temporal and spatial dimensions through multi-channel crawling. The data include macroeconomic factors such as Gross

Domestic Product, Consumer Price Index, Per Capita Disposable Income, and residential commercial properties; architectural features including building age, decoration level, and number of rooms; and neighborhood characteristics such as proximity to educational and medical facilities, transportation, and amenities within a 15-minute walk. Due to the dense distribution of data points and variable collinearity, the Bootstrap algorithm was employed to sample 1,000 iterations. After evaluating several models -General Linear Regression, Geographically Weighted Regression, Geographically Weighted Ridge Regression, and Geographically Weighted Lasso Regression- the latter was chosen for its superior data fit and minimal error margin, yielding significant insights.

(1) Spatial Impact: There is a notable effect of the spatial dimension on the prices of second-hand green housing, with variable impacts across different regions. From a broad distribution perspective, the prevalence of star-rated second-hand green housing diminishes from coastal to inland areas, such as from the Bohai Economic Rim and the Yangtze River Delta Economic Zone. Generally, higher star ratings positively influence housing prices. Macroeconomic conditions also affect these prices; residential commercial properties, CPI, and Gross Domestic Product generally have a positive impact, whereas Per Capita Disposable Income shows a negative correlation. Architectural features and neighborhood characteristics generally exhibit a negative influence. The standard deviation indicates the largest variability in Per Capita Disposable Income nationwide, while star ratings show the least variability (smallest data fluctuations). From a median perspective, at a national intermediate level, only architectural features and the Consumer Price Index negatively impact housing prices, with all other variables showing positive effects. From a max-min perspective, housing prices in specific areas significantly respond to provincial macroeconomic factors.

(2) Star Ratings: Nationally, higher overall star levels correspond with increased housing prices, though geographical variances exist. Among 71 communities, the star rating coefficient predominantly clusters around 0.1 ± 0.05, reflecting generally positive effects on second-hand housing prices. In the southeastern coastal regions, higher star ratings correlate with higher housing prices, while in central and northeastern regions, star ratings negatively affect housing prices.

(3) Macroeconomic Perspective: Across various regions, key macroeconomic variables such as Gross Domestic Product, Per Capita Disposable Income, and residential commercial properties generally have a positive impact on housing prices, whereas the Consumer Price Index often negatively affects them. Areas with higher Gross Domestic Product rankings experience less impact on housing prices. The Consumer Price Index commonly shows a negative distribution across regions; excluding housing consumption expenditure, increased consumption reduces assets available for property acquisition, resulting in a negative correlation. Conversely, a positive distribution in Per Capita Disposable Income suggests that higher disposable income and subsequent long-term asset accumulation positively influence housing prices. The correlation between market prosperity and housing prices is weaker in northern regions, whereas a thriving real estate market in other regions correlates with higher second-hand housing prices.

(4) Architectural and Neighborhood Characteristics: Most coefficients for architectural features are concentrated around 0 ± 0.1. In Northeast and Southwest China, architectural features negatively impact housing prices, while in Central and East Coastal China, they positively influence second-hand house prices. For neighborhood characteristics, most coefficients cluster around 0 ± 0.2. In southeast coastal areas, superior neighborhood characteristics correlate with lower second-hand house prices, whereas in central and northeastern areas, better neighborhood characteristics are associated with higher house prices.

Based on the above analysis, house prices show a negative relationship with star ratings, Gross Domestic Product, and other factors, deviating from traditional perceptions. From the perspective of star ratings, green-starred housing typically commands a green premium upon its initial sale (Zhang et al., 2023b). Traditionally, this premium decreases in the second-hand market but still exists. However, in the second-hand housing market, inadequate promotion (Zhang et al., 2021) and property depreciation can diminish the positive impact of star ratings. Additionally, macroeconomic factors influence green consumption behaviors (Ritter et al., 2015). In economically less developed areas, green star ratings do not significantly increase second-hand housing prices, as residents often prioritize basic housing functions.

From a macroeconomic perspective, the impact of macroeconomic factors on housing prices in some regions is negative, possibly due to temporal lags. Although lagged variables were incorporated for data analysis, the effects of time are not fully accounted for (Zhang et al., 2017). Furthermore, consumers purchasing second-hand green housing often opt for payment methods such as long-term savings or loans (Lee and Yang, 2019). The prosperity level of the real estate market in a given year tends to influence the initial sale of properties more directly (Glaeser et al., 2017), resulting in a mix of positive and negative effects among various macroeconomic variables.

From the perspective of architectural features, variables such as the orientation of the house and elevator availability, which are basic amenities of green housing, were excluded in the initial screening. The analysis focused on the number of rooms, level of decoration, and age of the house. Despite this focus, a negative impact was still observed. This is attributed to green housing with star ratings typically being sold with high-quality finishes to ensure its green and healthy development ethos (Coombs et al., 2016). Additionally, the issue of housing being renovated during multiple resales means that the age of a house does not solely determine its price negatively. Research by Yuan et al. (2022) found that while older green housing tends to be priced lower, some newer green housings increase in price as they age. This contradicts findings by scholars Clark and Lomax, who observed a minimal impact from the number of rooms since the basic layout of houses remains consistent.

From the standpoint of neighborhood characteristics, the impact of this indicator on housing prices is negative in some regions, which can be attributed to the potential negative effects of transport accessibility. While closer transport distances enhance convenience, they may lead to congestion in some cities, thereby affecting the residential living experience (Tang, 2021). Additionally, there is significant variability in transport conditions between regions. Jin et al. (2022) found that the impact of different transport distances on housing prices varies by region. Accessibility to transportation has a positive effect on housing prices, but negative impacts may emerge over time. Real estate developers, when choosing locations for green residential developments, consider both comfort (avoiding congestion and noisy transport routes) and convenience (proximity to transportation).

6.2 Policy implications.

Based on empirical analysis, the following policy recommendations are proposed as valuable references for policy formulation and practice:

(1) Policy Differentiation: Due to the significant impact of spatial dimensions on second-hand green housing prices, it is advisable for governments to account for regional differences when crafting policies. This avoids a one-size-fits-all approach and achieves more targeted policy orientations. In southeastern coastal regions, where the star rating of green housing positively correlates with housing prices, it is recommended to encourage developers to increase the star levels of residences through certifications, tax incentives, and other measures, thereby enhancing their market value. Conversely, in central and northeastern regions, where star ratings negatively impact housing prices due to limited market recognition or differing consumer preferences, governments should focus on targeted publicity, educational initiatives, and market promotion strategies to boost the appeal of star-rated housing.

(2) Macroeconomic Indicators and Housing Price Dynamics: Given the positive correlation of Gross Domestic Product, Per Capita Disposable Income, and residential commercial properties with housing prices, governments should monitor these economic indicators closely. This is particularly important in economically robust regions to mitigate the risk of market bubbles caused by rapid price increases. Conversely, considering the inverse relationship between the Consumer Price Index and housing prices, housing price control policies should be crafted with consumer spending pressures in mind to prevent adverse effects on residents' quality of life due to escalating housing costs.

(3) Strengthening Supervision and Assessment of Architectural Features and Neighborhood Characteristics: Governments should enforce stricter standards for building quality and design, intensifying oversight of architectural features to ensure the quality of housing in the second-hand market. Regular evaluations of neighborhood characteristics, including infrastructure completeness and public services, should be conducted to provide homebuyers with comprehensive and accurate information and to encourage developers to enhance these attributes. Urban planning should be refined to be more rational and scientific, with strategies tailored to local conditions to optimize the housing supply structure. In the development of new urban areas, a focus should be placed on enhancing public facilities and support services to improve residents’ convenience.

(4) Enhancing Data Collection and Sharing: Governments should establish and refine a real estate data collection mechanism, incorporating digital applications more effectively into the second-hand housing transaction sector to ensure data accuracy and timeliness, particularly for information on star-rated green second-hand housing. A unified information platform should be developed to aid homebuyers in searching and

comparing property information, thus increasing market transparency and fairness. Additionally, interdepartmental cooperation should be strengthened to facilitate data sharing and enable a more comprehensive understanding of real estate market dynamics.

6.3 Limitations and future research directions.

This study is limited by the imperfect disclosure of green building star ratings in the existing second-hand housing market, and the scope of the data requires further expansion due to disclosure limitations. Future research should standardize the house prices of second-hand green housing across different countries for comparison, to explore the variability in price influencing factors under different green standard systems. This approach will facilitate empirical contributions to the global promotion of green housing. Subsequent studies will continue to include policy factors as explanatory or moderating variables in the model to assess their impact on the prices of second-hand green housing. By examining the differences in policy environments, such as incentives for green building and enforcement of environmental regulations across various countries, the research will elucidate the role of these policies in shaping market supply and demand dynamics, thereby influencing green housing prices.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements:

This study was financially supported by grants from the National Natural Science Foundation of China (Nos. 72204098,72074186), the Major Project of Philosophy and Social Science Research in Colleges and Universities in

Jiangsu Province(No. 2022SJZD147), the Youth Project of Social Science Foundation of Jiangsu Province (No. 22GLC018), the Major Program of the National Fund of Philosophy and Social Science of China (No. 21&ZD109), Basic Scientific Center Project of National Science Foundation of China (No.71988101), the Fundamental Research Funds for the Central Universities concerned Chinese modernization (No. 20720231061), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX24_2489).

CRediT authorship contribution statement

Qianwen Li: Formal analysis, Conceptualization, Methodology, Writing-original draft. Tingyu Qian: Data curation, Writing-original draft, Software，Methodology, Visualization. Hui Wang: Methodology, Visualization, Software， Writing – review & editing . Chuanwang Sun: Resources, Supervision, Writing – review & editing， Visualization.

References

André, C., Gabauer, D., & Gupta, R. (2021). Time-varying spillovers between housing sentiment and housing market in the United States. Finance Research Letters, 42, 101925.

Agatz, N., Fan, Y., & Stam, D. (2021). The impact of green labels on time slot choice and operational sustainability. Production and Operations Management, 30(7), 2285-2303.

Aslett, K., Sanderson, Z., Godel, W. et al. Online searches to evaluate misinformation can increase its perceived veracity. Nature 625, 548 – 556 (2024).

Bednarek, P., Kaat, D. M. T., Ma, C., & Rebucci, A. (2021). Capital flows, real estate, and local cycles: Evidence from German cities, banks, and firms. The Review of Financial Studies, 34(10), 5077-5134.

Bedendo, M., Nocera, G., & Siming, L. (2023). Greening the financial sector: Evidence from bank green bonds. Journal of Business Ethics, 188(2), 259-279.

Chen, Y., Shi, S., & Tang, Y. (2019). Valuing the urban hukou in China: Evidence from

a regression discontinuity design for housing prices. Journal of Development Economics, 141, 102381.
Chen, K., Lin, H., Cao, F., Han, Y., You, S., Shyr, O., ... & Huang, X. (2022). Do hospital and rail accessibility have a consistent influence on housing prices? Empirical evidence from China. Frontiers in Environmental Science, 10, 1044600.
Chen, S., Xiong, L., Ma, Q., Kim, J. S., Chen, J., & Xu, C. Y. (2020). Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method. Journal of Hydrology, 589, 125156.
Chen, Y., Jiao, J., & Farahi, A. (2023). Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit, and density factors on single-family housing price. Cities, 140, 104432.
Clark, S. D., & Lomax, N. (2018). A mass-market appraisal of the English housing rental market using a diverse range of modelling techniques. Journal of big Data, 5, 1-21.
Coombs, K. C., Chew, G. L., Schaffer, C., Ryan, P. H., Brokamp, C., Grinshpun, S. A., ... & Reponen, T. (2016). Indoor air quality in green-renovated vs. non-green low-income homes of children living in a temperate region of US (Ohio). Science of the Total Environment, 554, 178-185.
China Real Estate News, Youbo Group, Youlv Think Tank, & Lemon Tree. (2020). China Green Real Estate Development Report.
Dell’Anna, F., & Bottero, M. (2021). Green premium in buildings: Evidence from the real estate market of Singapore. Journal of Cleaner Production, 286, 125327.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2009). Geographically weighted regression. The Sage handbook of spatial analysis, 1, 243-254.
Glaeser, E., Huang, W., Ma, Y., & Shleifer, A. (2017). A real estate boom with Chinese characteristics. Journal of Economic Perspectives, 31(1), 93-116.
He, Y., Zhao, Y., & Tsui, K. L. (2021). An adapted geographically weighted Lasso (Ada-GWL) model for predicting subway ridership. Transportation, 48(3), 1185-1216.
Hui, E. C. M., & Yu, K. H. (2021). Housing market segmentation and the price effect of certified green residential properties. Habitat International, 111, 102350.
Huang, M. Y. (2023). Analyzing the effects of green building on housing prices: case study of Kaohsiung, Taiwan. Environment, Development and Sustainability, 25(2), 1205-1235.
Hsu, Y. L., & Chow, E. H. (2013). The house money effect on investment risk taking: Evidence from Taiwan. Pacific-Basin Finance Journal, 21(1), 1102-1115.
Huang, B., He, X., Xu, L., & Zhu, Y. (2020). Elite school designation and housing prices-quasi-experimental evidence from Beijing, China☆ Journal of Housing Economics, 50, 101730.
Hudson, R. (2024). Biodiversity and risk in the financial sector. Journal of Sustainable Finance and Accounting, 2, 100009.
Islam, H., Bhuiyan, M., Tushar, Q., Navaratnam, S., & Zhang, G. (2022). Effect of star rating improvement of residential buildings on life cycle environmental impacts and costs. Buildings, 12(10), 1605.
Jiang, Y., Xing, Y., Zhao, D., & Jiao, R. (2021). Resale of green housing compensates for its premium pricing: An empirical study of China. Journal of Green Building, 16(4), 45-61.
Jin, B., Geng, J., Ke, S., & Pan, H. (2023). Analysis of spatial variation of street landscape greening and influencing factors: an example from Fuzhou city,

China. Scientific Reports, 13(1), 21767.
Jin, T., Cheng, L., Liu, Z., Cao, J., Huang, H., & Witlox, F. (2022). Nonlinear public transit accessibility effects on housing prices: Heterogeneity across price segments. Transport Policy, 117, 48-59.
Jia, S., Wang, Y., & Fan, G. Z. (2018). Home-purchase limits and housing prices: Evidence from China. The Journal of Real Estate Finance and Economics, 56, 386-409.
Yoshida, J. (2020). The economic depreciation of real estate: Cross-sectional variations and their return implications. Pacific-Basin Finance Journal, 61, 101290.
Kleindorfer, P. R., Singhal, K., & Van Wassenhove, L. N. (2005). Sustainable operations management. Production and operations management, 14(4), 482-492.
Lee, Y., & Yang, K. (2019). Modeling diversification and spillovers of loan portfolios' losses by LHP approximation and copula. International Review of Financial Analysis, 66, 101374.
Li, H., Chen, P., & Grant, R. (2021a). Built environment, special economic zone, and housing prices in Shenzhen, China. Applied Geography, 129, 102429.
Li, H., Wei, Y. D., Yu, Z., & Tian, G. (2016). Amenity, accessibility and housing values in metropolitan USA: A study of Salt Lake County, Utah. Cities, 59, 113-125.
Li, W., Fang, G., & Yang, L. (2021b). The effect of LEED certification on office rental values in China. Sustainable Energy Technologies and Assessments, 45, 101182.
Li, J., Yuan, J., Huang, Y., & Zhang, Y. (2023). Revealing the widespread existence and serious adverse health consequences of low-price rental housing in urban villages in Xiamen, China. Humanities and Social Sciences Communications, 10(1), 1-11.
Liu, C., Wu, X., & Wang, L. (2019). Analysis on land ecological security change and affect factors using RS and GWR in the Danjiangkou Reservoir area, China. Applied Geography, 105, 1-14.
Lucey, B., Urquhart, A., & Zhang, H. (2022). UK Vice Chancellor compensation: Do they get what they deserve?. The British Accounting Review, 54(4), 101108.
Murakami, D., Tsutsumida, N., Yoshida, T., Nakaya, T., Lu, B., & Harris, P. (2023). A linearization for stable and fast geographically weighted Poisson regression. International Journal of Geographical Information Science, 1-22.
Naderi, S., Heslop, S., Chen, D., MacGill, I., & Pignatta, G. (2022). Consumer cost savings, improved thermal comfort, and reduced peak air conditioning demand through pre-cooling in Australian housing. Energy and Buildings, 271, 112172.
Naderi, S., Heslop, S., Chen, D., Watts, S., MacGill, I., Pignatta, G., & Sproul, A. (2023). Clustering based analysis of residential duck curve mitigation through solar pre-cooling: A case study of Australian housing stock. Renewable Energy, 216, 119064.
Nguyen, D. D., Ongena, S., Qi, S., & Sila, V. (2022). Climate change risk and the cost of mortgage credit. Review of Finance, 26(6), 1509-1549.
Nazemi, A., & Fabozzi, F. J. (2018). Macroeconomic variable selection for creditor recovery rates. Journal of Banking & Finance, 89, 14-25.
Porumb, V. A., Maier, G., & Anghel, I. (2020). The impact of building location on green certification price premiums: Evidence from three European countries. Journal of Cleaner Production, 272, 122080.
Pourmohammadi, P., Strager, M. P., Dougherty, M. J., & Adjeroh, D. A. (2021). Analysis of land development drivers using Geographically Weighted Ridge Regression. Remote Sensing, 13(7), 1307.
Peng, Y., Tian, C., & Wen, H. (2021). How does school district adjustment affect

housing prices: An empirical investigation from Hangzhou, China. China Economic Review, 69, 101683.

Qu, S., Hu, S., Li, W., Zhang, C., Li, Q., & Wang, H. (2020). Temporal variation in the effects of impact factors on residential land prices. Applied Geography, 114, 102124.

Ritter, Á. M., Borchardt, M., Vaccaro, G. L., Pereira, G. M., & Almeida, F. (2015). Motivations for promoting the consumption of green products in an emerging country: exploring attitudes of Brazilian consumers. Journal of Cleaner Production, 106, 507-520.

Safarova, S., Van Hoof, J., Law, L., Zander, K. K., & Garnett, S. T. (2022). Thermal comfort in a tropical savanna climate: The case of home occupants in Darwin, Australia. Energy and Buildings, 266, 112074.

Shen, Q., Pan, Y., & Feng, Y. (2023). The impacts of high-speed railway on environmental sustainability: quasi-experimental evidence from China. Humanities and Social Sciences Communications, 10(1), 1-19.

Sterbenk, Y., Champlin, S., Windels, K., & Shelton, S. (2022). Is femvertising the new greenwashing? Examining corporate commitment to gender equality. Journal of Business Ethics, 177(3), 491-505.

Tang, C. K. (2021). The cost of traffic: evidence from the London congestion charge. Journal of Urban Economics, 121, 103302.

Tasyurek, M., & Celik, M. (2022). 4D-GWR: geographically, altitudinal, and temporally weighted regression. Neural Computing and Applications, 34(17), 14777-14791.

Urolagin, S., Sharma, N., & Datta, T. K. (2021). A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. Energy, 231, 120963.

van der Drift, R., de Haan, J., & Boelhouwer, P. (2023). Mortgage credit and house prices: The housing market equilibrium revisited. Economic Modelling, 120, 106136.

Wang, J., & Li, G. (2022). Pursuing educational equality and divergence in the housing market: How do educational equality policies affect housing prices in Shanghai?. Cities, 131, 104001.

WANG, Qi, et al. The impacts of road traffic on urban air quality in Jinan based GWR and remote sensing. Scientific reports, 2021, 11.1: 15512.

Wu, D. (2020). Spatially and temporally varying relationships between ecological footprint and influencing factors in China’s provinces Using Geographically Weighted Regression (GWR). Journal of Cleaner Production, 261, 121089.

Xu, S. (2023). China’s climate governance for carbon neutrality: regulatory gaps and the ways forward. Humanities and Social Sciences Communications, 10(1), 1-10.

Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., ... & Zhang, N. (2021). Spatial downscaling of land surface temperature based on a multi-factor geographically weighted machine learning model. Remote Sensing, 13(6), 1186.

Yang, L., Chu, X., Gou, Z., Yang, H., Lu, Y., & Huang, W. (2020). Accessibility and proximity effects of bus rapid transit on housing prices: Heterogeneity across price quantiles and space. Journal of Transport Geography, 88, 102850.

Yang, P., Verhoef, W., Prikaziuk, E., & van der Tol, C. (2021). Improved retrieval of land surface biophysical variables from time series of Sentinel-3 OLCI TOA spectral observations by considering the temporal autocorrelation of surface and atmospheric properties. Remote sensing of environment, 256, 112328.

Yii, K. J., Tan, C. T., Ho, W. K., Kwan, X. H., Nerissa, F. T. S., Tan, Y. Y., & Wong,

K. H. (2022). Land availability and housing price in China: Empirical evidence from nonlinear autoregressive distributed lag (NARDL). Land Use Policy, 113, 105888.

Yuan, Z., Zhao, M., & Du, M. (2022). Price discount of green residential building in China’s second-hand housing market: Evidence from Shenzhen. Sustainable Energy Technologies and Assessments, 52, 102171.

Zhang, B., Zhang, Y., Li, J., Song, Y., & Wang, Z. (2023b). Does the energy efficiency of buildings bring price premiums? Evidence from urban micro-level energy data. Renewable and Sustainable Energy Reviews, 181, 113315.

Zhang, F., Zhang, H., & Zhang, Y. (2023a). Trust premium in the second-hand housing market: evidence from the negotiation rate. Humanities and Social Sciences Communications, 10(1), 1-15.

Zhang, X., Lin, Z., Zhang, Y., Zheng, Y., & Zhang, J. (2021). Online property brokerage platform and prices of second-hand houses: Evidence from Lianjia’s entry. Electronic Commerce Research and Applications, 50, 101104.

Zhang, W., Lu, D., Zhao, Y., Luo, X., & Yin, J. (2022). Incorporating polycentric development and neighborhood life-circle planning for reducing driving in Beijing: Nonlinear and threshold analysis. Cities, 121, 103488.

Zhang, H., Hudson, R., Metcalf, H., & Manahov, V. (2015). Identification of house price bubbles using user cost in a state space model. Applied economics, 47(56), 6088-6101.

Zhang, H., Hudson, R., Metcalf, H., & Manahov, V. (2017). Investigation of institutional changes in the UK housing market using structural break tests and time-varying parameter models. Empirical economics, 53, 617-640.

Zheng, H., Qian, J., Liu, G., Wu, Y., Delang, C. O., & He, H. (2023). Housing prices and household consumption: a threshold effect model analysis in central and western China. Humanities and Social Sciences Communications, 10(1), 1-16.

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?. Journal of Empirical Finance, 54, 97-117.

Highlights:

- Explores macro and micro factors influencing green housing prices regionally.
- Analyzes 20 years of data using big data multi-channel scraping techniques.
- 3Combines Bootstrap and Lasso regression to handle col linearity and clustering.
- Incorporates GWR model for region-specific analysis of housing price dynamics.
- Enhances predictive accuracy and stability for price factor analysis.