Hedonic House Prices and Spatial Quantile Regression

Wen-Chi Liao

Xizhu Wang

July 2010
Hedonic House Prices and Spatial Quantile Regression

Wen-Chi Liaoa,1; Xizhu Wangb

aN, bNational University of Singapore

July 2010

1Department of Real Estate, National University of Singapore. 4 Architecture Drive, 117566 Singapore. E-mail: wliao@nus.edu.sg; Fax: +65-6774-8684. We thank Yuming Fu, Hongyu Liu, and Yong Tu for comments and feedback. All errors are ours.
Abstract

Despite its long history, hedonic pricing for housing valuation remains an active research area, and applications of new estimation methods continually push research frontiers. However, housing studies regarding Chinese cities are limited due to the short history of China’s free housing market. Such studies may, nonetheless, provide new insights given the nation’s current transitional stage of economic development. Therefore, in this research, we utilize publicly accessible sources to construct a new dataset for an emerging Chinese city—Changsha, and incorporate quantile regression and spatial econometric modeling to examine how implicit prices of housing characteristics may vary across the conditional distribution of house prices. Substantial variation exists across quantiles, suggesting that ordinary regression is insufficient on its own. Quantile estimates of a spatial-lag model show considerable spatial dependence in the upper and lower parts of the distribution but little dependence in the medium range. Several other interesting patterns are also found, and their intuitions and implications are discussed.
1 Introduction

Despite its long history, hedonic pricing for housing valuation remains an active research area, and applications of new estimation methods continually push research frontiers. However, housing studies regarding Chinese cities are limited due to the short history of China’s free housing market. Such studies may, nonetheless, provide new insights given the nation’s current transitional stage of economic development. Therefore, in this research, we utilize publicly accessible sources to construct a new dataset for an emerging Chinese city—Changsha, and incorporate quantile regression and spatial econometric modeling to examine how implicit prices of housing characteristics, accessibility to greenery spaces and city center in particular, vary across the conditional distribution of house prices. The findings are intuitive and have interesting implications.

The origin of hedonic pricing method can be dated back to Hass (1922) and Court (1939), while its theoretical foundation is provided by Lancaster (1966) and Rosen (1974). Since Ridker and Henning (1967), hedonic pricing models have been widely used in housing research and appraisals. The widespread popularity is largely due to the fact that hedonic pricing method is a revealed preference approach allowing housing expenditure to be decomposed into the values and quantities of individual housing characteristics. Comprehensive reviews can be found in Follain and Jimenez (1985), Sheppard (1999), Malpezzi (2003) and Sirmans, Macpherson, and Zietz (2005).

Although housing studies usually adopt hedonic regression models, their coefficient estimates for certain housing characteristics often differ in the magnitude, level of significance, and sign. Sirmans, Macpherson, and Zietz (2005) review the results of about 125 recent empirical studies. They find that such differences exist even for common characteristics in hedonic pricing equations. For instance, 15 studies identify the effects of housing units’ distance to CBD on prices, 5 of these studies report a negative impact, while another 5 studies show a positive impact. The remaining 5 studies illustrate an insignificant effect.

There are various reasons for the diverse results in the literature. The most obvious yet perhaps the least worrisome is the specificity of the results to its market of study. For example, people in a small town may value the distance to CBD differently from the residents in a big city. Similarly, Americans may value certain housing characteristics in a different light from the Chinese. Hence, the results obtained from one study cannot be applied universally to all housing markets.

Spatial dependence—often used interchangeably with spatial autocorrelation—is another well appreciated reason that may account for the different and sometimes inconsistent results found in the hedonic
house price literature. It occurs when an observation at one location depends upon observations at other locations. It can be resulted from the spillover effects of prices of nearby housing properties, omitted spatially correlated variables, as well as measurement error and model misspecification. The presence of spatial dependence is detrimental to the efficiency and unbiasedness of OLS estimators in traditional hedonic models, as issues of autocorrelation and endogeneity can prevail. Since spatial dependence often occurs among housing properties, it is a caveat in using hedonic pricing model for housing research. Developments of spatial econometrics have addressed these issues and provided various remedies. Anselin (1998) and LeSage and Pace (2009) offer a wide coverage of these methods.

A third reason for the variations in past hedonic price estimates is that housing characteristics may be valued differently at different points of the conditional distribution of house prices, which is referred to as quantile effects in this paper. The literature had not examined this dimension until recent applications of quantile regressions. McMillen and Coulson (2007) and McMillen (2008), who construct quantile house price indexes and study house price appreciation, identify significant variations in values of physical attributes across quantiles. Mak, Choy, and Ho (forthcoming) find that quantile effects exist even in one single condominium. Zietz, Zietz, and Sirmans (2008) apply a new development of spatial quantile regression and use multiple-listing-service data of Orem-Provo, Utah, United States. They find that some variables, such as square footage, lot size, bathrooms and floor type, have a greater impact on higher-priced homes. However, they observe negligible spatial dependence and conclude greater importance of examining quantile effects. Also, their quantile estimates for the distance to the center of Orem are all positive but mostly insignificant and do not demonstrate a clear pattern of change. In addition, they do not consider the accessibility to public green spaces. These observations are likely due to the fact that the subject examined is the housing market of a small city in Utah, a state with mild income inequality as well as rather homogeneous demographics.

Existence of quantile effects may partly explain the variations in past hedonic price estimates obtained through mean regression. A plausible reasoning is articulated in this and the next paragraph. Housing researchers have well appreciated that heterogeneous households, such as the rich and the poor, may value housing characteristics differently (see Malpezzi, 2003), and affirmative empirical evidences are found in Bayer, McMillan and Rueben (2004). In the empirical part of this paper, which concerns equilibrium residential sorting, the authors make use of the restricted-access microdata of U.S. Census of Population and Housing. Since every microdatum is identifiable at the level of census blocks (very fine geographic units), they are able to produce estimates that give a precise characterization of preferences. One of their findings is that marginal willingness to pay for desirable housing characteristics and location attributes,
which include neighborhood socio-demographic compositions, accessibility to workplaces, and proximity to various amenities, increases with income.

Unfortunately, a dataset like the restricted-access U.S. Census is rare, and access to that data source is a privilege. Usually, many influential attributes are unobservable in the housing studies that apply hedonic pricing methods. This implies that the response distribution could depend upon covariates and contain quantile effects. For example, if the rich consistently place a higher value on unobservable characteristics than the poor, then housing units at upper quantiles of the response distribution would be occupied by the rich in the sorting equilibrium. This underlines the common statement in the literature that higher income households buy higher-priced homes. As a result, implicit prices of observables at the upper quantiles could be driven by the willingness to pay of the rich. If the rich adore a particular observable housing characteristic while the poor dislike it, then the implicit prices of the characteristic at the upper and lower parts of the response distribution would exhibit opposite signs—quantile effects. In this situation, mean regression may not provide useful information on this particular housing characteristic, as the implicit price that the regression produces can be positive, negative, or insignificant.

Although applications of hedonic pricing models in housing research are widespread and many of them use cutting-edge methods, studies on Chinese cities are still limited (see Chen and Hao, 2008, for a discussion on this matter) due to restricted data availability for housing transactions, as China’s free housing market exists for only 10 years. The State used to order work units (plants) to provide housing for their employees at a very low and subsidized rent. The employees received housing arrangements based on their rank and seniority with no right to choose. The reform in 1998 ceased this welfare allocation of housing. Market prices have prevailed since then. Many cities currently have multiple listing services for housing properties, but the data holdings are usually not detailed and lack essential information for research purposes. The data limitation prevents knowledge accumulation on valuation of housing in China.

As China is in a unique transitional stage of economic development, studies of hedonic house prices for Chinese cities can add new insights to the literature and offer useful policy implications. Furthermore, applications of quantile regression may be particularly interesting and helpful. To elaborate this point, the households’ willingness to pay for green space is used as an illustration. There might be a stereotypical image that, in China, insufficient attention is given to the greening of the environment. However, demand for quality of urban living, which is a normal good, should have increased substantially, as China has achieved impressive urban economic development. Indeed, Zheng, Fu and Liu (2009) find supporting evidences. Nevertheless, China is still a developing country and the spread of urban households’ income...
distribution is broad (Zenou 2010). Thus, accessibility to green space could be valued quite differently in the upper and lower parts of the response distribution.

The Changsha city was chosen for this study for the following reasons. Firstly, the city’s micro data that are available from unrestricted public domains are relatively detailed, and more importantly, Changsha is an emerging, transitional city. The city has shown astonishing economic growth, despite its slow development during the early phase of Reform and Opening Up. Changsha’s average annual GDP growth was 15.4% from 2003 to 2007, compared with the national average of 11% during the same period. In 2007, GDP per capita in the urbanized areas of Changsha was $7,450 USD, which was up to 78% of that of Shanghai\(^1\). As China is shifting its development focus to its Middle Region, the future of Changsha, the richest city in this region\(^2\), is promising. Furthermore, Changsha has great emphasis and immense spending on its greening and transportation infrastructure. This allows the results derived in this paper to shed some light on government policies.

In this research, quantile regression is used to estimate a spatial-lag model. The results suggest existence of substantial quantile effects, especially for housing characteristics like the proximity to the CBD and availability of green spaces. We also discover considerable positive spatial dependence in the upper and lower parts of the conditional distribution of house prices but little dependence in the medium range. This finding appeals to examining quantile effects together with spatial dependence.

The rest of this paper is organized as follows. Section 2 lays out the research methodology, including brief reviews on quantile regression and spatial-lag model and ways to integrate them. Section 3 introduces our data, including discussions on related issues. In section 4, we present estimation results and discuss their economic intuitions and implications. Section 5 concludes.

### 2 Methodology

As it is aimed to incorporate quantile regression into spatial econometric modeling to examine quantile effects in Changsha’s housing market, this section first reviews quantile regression and spatial econometrics and then introduces the methods to integrate the two disciplines of econometrics. It also includes specific discussion on our estimation method.

Unlike classical linear regression that estimates a conditional mean function, quantile regression estimates a conditional quantile function, in which a quantile of the response variable’s conditional dis-
tribution is expressed as a function of covariates. Since quantile regression allows its estimates to vary with the corresponding quantile, it is particularly useful when quantile effects exist. Quantile regression is preferred to the alternative approach that first subdivides the sample according to the unconditional distribution of the response variable and subsequently performs OLS for each subsample. This is because quantile regression uses the full sample and avoids the truncation problem that the alternative approach usually encounters. Other important advantages of quantile regression include its superior capability in handling heteroscedasticity, outliers, and unobserved heterogeneity. Koenker and Hallock (2001) and Koenker (2005) review this econometric method thoroughly.

The mechanism to carry out quantile regression is similar to ordinary regression. The difference is, instead of searching for the argmin of sums of squared residuals, quantile regression looks for the argmin of weighted sums of absolute residuals. Consider a sample of $N$ observations for the estimation of a hedonic house price model. Included in the sample are $p$ and $X$, where $p$ denotes an $N$ vector of house prices, and $X$ denotes an $N \times K$ matrix in which the first column is a column of ones and the rest record the values of $K - 1$ explanatory variables. The minimization problem of a quantile regression can be written as:

$$\hat{\beta}_q = \arg\min_{\beta_q \in \mathbb{R}^K} \sum_{n=1}^{N} |p_n - x_n \beta_q| \omega_n$$

where $\hat{\beta}_q$ is the vector of coefficient estimates, and the subscript $q \in (0, 1)$ denotes the quantile to be estimated. In addition, $p_n$ is the $n$-th entry of $p$, $x_n$ is the $n$-th row of $X$, and $\omega_n$ is the $n$-th observation's weight which is defined as:

$$\omega_n = 2q$$

if $p_n - x_n \beta_q > 0$, and

$$\omega_n = 2 - 2q$$

otherwise. As for the standard errors of the coefficient estimates, they can be estimated using bootstrapping.

Housing property prices often involve positive spatial autocorrelation, which means that prices of geographically nearby units tend to be similar due to spatial dependence of properties. When the data generating process encounters spatial dependence, issues of efficiency and unbiasedness of OLS estimators arise. Spatial econometric modeling therefore becomes necessary. One common approach is to model spatial dependence through a spatial weight matrix, and housing studies often incorporate the weight matrix into one of the two alternative econometric models: spatial-lag model and spatial error model.
(As discussed in Kim, Phipps and Anselin, 2003, the two models are related mathematically but are interpreted differently.)

For the spatial-lag model, its general form can be written as:

\[ p = \lambda W p + X \beta + \varepsilon \]

where \( W \) is an \( N \times N \) spatial weight matrix, and \( \lambda \) indicates the degree of spatial autocorrelation and is a parameter to be estimated. For the system to be stationary, \( 1 \geq \hat{\lambda} \geq -1 \) must hold. If \( \hat{\lambda} > 0 \), spatial autocorrelation is positive, and prices of nearby properties tend to be similar. In other words, the pattern of spatial clustering of similarly priced houses is more pronounced than randomness. On the other hand, if \( \hat{\lambda} < 0 \), spatial autocorrelation is negative, and prices of nearby properties tend to be dissimilar.

In the above model, the spatial weigh matrix \( W \) defines how a property is affected by nearby properties. The literature has documented various types of specifications that can be broadly classified into contiguity and distance-based matrices. The matrix is often row-standardized to constrain the sum of elements of each row to be equal to 1, so that the above model conceptually means that the price of each property is affected by a form of weighted average prices of nearby properties.

If the above model correctly specifies the data generation process, then OLS is inappropriate for the estimation. First, the model indicates autocorrelation in the data, which violates the classical linear regression’s assumption of no correlation in disturbances. With the presence of autocorrelation, OLS estimators are inefficient and their variances are biased. Furthermore, the above model is in fact a simultaneous system of equations that encounters endogeneity and leads to correlation between the spatially lagged variable and the error term. With endogeneity, OLS estimators are not even consistent.

The alternative model is the spatial error model, whose general form can be written as:

\[ p = X \beta + \mu \]
\[ \mu = \lambda W \mu + \varepsilon \]

Characterized by a spatially lagged error term, housing researchers who adopt this model often assume that there are omitted variables that are spatially dependent. This interpretation is different from that of the spatial-lag model, which hypothesizes spatial spillovers of prices. If the spatial error model correctly describes the data generation process, then OLS estimators are inefficient although unbiased.

This research adopts a spatial-lag model to tackle spatial autocorrelation, for three reasons. First
and foremost, quantile regression is not yet applicable to estimation of a spatial error model. Second, the prevalent form of residential development in China is condominium. The substantial clustering of housing units at a single location makes a spatial error model less suitable. Last but not least, a spatial-lag model can exploit our large sample of Changsha housing transactions and consistently estimate housing characteristics’ implicit prices.

To model the structure of the spatially lagged housing prices, we construct a spatial weight matrix using the following rules. Let \( d_{ij} \) denote the distance from development \( i \) to development \( j \) in kilometers. \( \forall i, j \in \{1, ..., N\} \), we have:

\[
w_{ij} = z_{ij} / s_i
\]

where

\[
z_{ij} = \begin{cases} 
  e^{-d_{ij}} & \text{if } d_{ij} \leq 4 \\
  0 & \text{if } d_{ij} > 4 
\end{cases}
\]

and

\[
s_i = \sum_{j=1}^{N} z_{ij}
\]

This distance decay matrix assigns nearby housing units a higher weight than those that are further away. Previously, we also considered using a binary distance matrix, in which non-diagonal elements are equal to 1 if \( d_{ij} \) is smaller than a cut-off value of distance and 0 otherwise. However, the resulted estimates, the spatial autocorrelation parameter in particular, were very sensitive to the cut-off value, due to the prevalence of condominiums\(^3\). Therefore, the above distance decay matrix is more favorable.

The matrix’s cut-off value is set to 4 kilometers. On the one hand, past research often uses a long cut-off distance to define neighborhoods (see, for instance, Dubin, 1998; and Kim, Phipps and Anselin, 2003). On the other hand, the longest distance between any pairs of developments in our sample is about 24 kilometers. After all, the effect of an increase or decrease from the 4-km cut-off on estimation results is minimal, given the construction of the weight matrix.

There are two methods to integrate quantile regression with spatial econometric modeling. The first one is Two Stage Quantile Regression (2SQR) proposed by Kim and Muller (2004), and the other is Instrumental Quantile Regression (IVQR) introduced by Chernozhukov and Hansen (2006). These two papers are, in fact, not specific to spatial econometrics but rather more concerned with endogeneity. Nevertheless, the methods introduced broaden the possibility of using quantile regression to estimate a spatial-lag model. Su and Yang (2008) have been working on a theatrical IVQR research that is more

\(^3\)The result can change considerably, when a perturbation in the cutoff value includes or excludes some large scale developments.
specific to spatial econometrics. Zietz, Zietz, and Sirmans (2008) deliver an empirical application of 2SQR. Kostov’s (2009) discussion compares 2SQR and IVQR. Basically, 2SQR is in line with 2SLS, and IVQR is asymptotically equivalent to GMM (Chernozhukov and Hansen, 2006). The advantage of 2SQR in estimating the spatial-lag model is its computational simplicity, as compared to IVQR. However, 2SQR relies on asymptotic properties, while IVQR has better finite sample properties. Given that we have a large sample, we choose 2SQR.

We estimate the following model:

\[ p = \lambda_q W p + X \beta_q + \varepsilon_q \]

where \( p \) is the log-transformed house prices. The subscript \( q \) indicates the corresponding quantile, and \( \lambda_q \) and \( \beta_q \) are to be estimated. The estimation method involves two stages. In the first stage, the spatially lagged endogenous variable \( W p \) is regressed against the spatially lagged exogenous variables \( W X \) as well as \( X \) for a quantile. The predicted \( \hat{W} p \), which is obtained from the first stage regression, is substituted for \( W p \) in the spatial-lag model to eliminate correlation between the spatially lagged endogenous variable and the error term. Then, the second stage regression for that quantile is performed to obtain \( \hat{\lambda}_q \) and \( \hat{\beta}_q \). This two-stage procedure is repeated for other quantiles.

3 Data

The data, available upon request, were manually collected from unrestricted public domains, as we had no access to any readily available dataset on Changsha’s housing market. Information on individual properties was gathered from Changsha Real Estate Information Center, a subsidiary body of Changsha Municipal Bureau of Property Right Management. The Center’s website\(^4\) provides records of new residential properties, sold or approved for sale, in Changsha city. All records contain essential information on the house price, floor area, floor number, greening rate\(^5\), number of bedrooms, plot ratio, sale approval date, sale status, and shared common area. The address of each property is also available to enable the identification of its longitude and latitude using Geographic Information System. The geographic coordinates allow users to calculate the distance from each property to the central business district, the nearest park, and all other properties, using the Haversine Formula of Great-circle Distance\(^6\).

\(^4\)http://www.0731fdc.com
\(^5\)The term, greening rate, is widely used in China. Simply put, it is the share of land area which is green in the satellite image.
\(^6\)The Great-circle Distance is defined as the shortest distance between two points of the surface of a sphere which are measured along a path on the sphere’s surface. As the Earth is an approximate sphere, formulas of the Great-circle Distance...
A caveat is that the Center does not provide the property-transaction date. The solution was to access the website in September 2009 and record properties that were already sold and were approved for sale after August 2008. Thus, the properties included in the sample were all transacted within this one-year period, during which Changsha’s housing prices were relatively stable. Furthermore, the sample only includes sales in Changsha’s urbanized areas. In total, the one-year sample includes 46,356 residential property sales in 113 residential developments.

Distance from each property to its nearest park is a variable included in the sample. To construct this variable, information from the Changsha Chorography and Almanac Office was used to identify a total number of 17 parks in the urbanized areas. Next, the geographic coordinates of each park’s center point were located for distance calculation. The minimum scale of these parks is 1.12 hectares. The city government has an additional 4 parks. However, they are outside the urbanized areas and are not the nearest park to any developments in the sample. Community gardens within private residential developments are also not included, as they are not available to the public. Their impact would be captured by another variable, the greening rate.

Following Lutzenhiser and Netusil (2001), the 17 parks are categorized into two types: urban parks and natural parks. Urban parks have more than 50% of land areas manicured or landscaped and developed for non-natural-resource-dependent recreation, such as playground and skating rink. Natural parks have more than 50% of land areas preserved for native or natural vegetation and the park use is balanced between preservation of natural habitat and natural-resource-based recreation, such as boating and camping. Two out of the 17 parks are natural parks.

Changsha’s urbanized areas consist of five municipal districts. In regression, it would be helpful to identify these district governments’ fixed effects, because district governments in China play important roles in managing and facilitating economic activities. District governments draw up social and economic development plans, enforce laws, provide public infrastructures and execute housing policies.

Figure 1 maps the five districts. In clockwise order, from the left, they are Yuelu, Kaifu, Furong, Yuhua and Tianxin. The furthermost north of Kaifu and south of Yuhua and Tianxin are not shown, as none of the developments in the sample is there. In addition, the map marks the CBD with a money sign and the residential developments, urban parks and natural parks with nail-, tree- and bushwalking-labels, respectively. The longest distance between any pair of developments on the map is about 24 kilometers.

[Insert Figure 1 here]

can be used to calculate the distance between any two points on the Earth’s surface. For our purpose, the ideal formula is the Haversine Formula, because it balances between accuracy and computation as compared with Arccosine and Vincenty Formulas.
Table 1 provides a complete description of variables that will be used in regression. There is no dummy variable to indicate whether a unit is in Tianxin district, as this is the base group for comparison. Table 2 gives the summary statistics for the variables.

Concerns over asymmetric information may not be a significant issue in this research, as the sample period starts from mid 2008. Previously, pre-sale properties, which are supplied to the market before project completion, were not protected by “no lemon laws.” Thus, Tse and Love (2000) and Yang (2001) argue that research using transaction data of new housing properties in China may be subject to asymmetric information, as pre-sale is the dominating form of sales. Nevertheless, in 2004, the Ministry of Housing and Urban-Rural Development of the People’s Republic of China amended the statute of Urban Residential Properties Pre-sale Management Method in Private Sector. The purpose was to reduce hidden forward risks involved in forward property market. The amendment sets unambiguous requirements for developers who wish to sell properties before project completion. The developers must pay Land-use-right Transfer Payment in full; obtain Land-use-right Certificate, Construction Project Planning Permit, and Construction Permit; invest at least 25% of total project costs; and confirm construction schedule and completion date, before applying for a compulsory Pre-sale Certificate.

4 Estimation Results

This section begins with a comparison between OLS and 2SLS estimates, which concern the conditional mean. Then, an examination of quantile estimates reveals that taking spatial dependence into account can considerably affect the estimators at certain quantiles and putting quantile regression into practice can quantify substantial variations in the implicit prices along the house price distribution. The quantile effects are also illustrated in quantile charts.

The OLS and 2SLS estimates are presented in the first column of Tables 3 and 4, respectively. The adjusted $R^2$ of both models are well above 0.8. The spatially lagged variable’s 2SLS coefficient estimate $\hat{\lambda}$ is 0.0986, which significantly indicates positive spatial autocorrelation of house prices. A housing unit in a more expensive neighborhood tends to be dearer than an otherwise identical unit in a less expensive neighborhood. Comparing the results of OLS and 2SLS, it can be noted that correcting spatial autocorrelation has greater impact on the coefficient estimates of project and location specific characteristics than the estimates of unit specific characteristics and the overall goodness of fit.
Many of the OLS and 2SLS coefficient estimates bear the same sign as the majority of 125 studies surveyed in Sirmans, Macpherson and Zietz (2005). House prices are expected to increase with floor area and numbers of bedrooms but decrease with the distance to CBD. Nevertheless, the estimates for Changsha do not always tally with those for western cities. For instance, a unit on a higher floor can command a higher price, but this result is not surprising given Asia’s high-density built environment.

The external share and plot ratio are not among the variables that commonly appear in western studies, but they are relevant to markets with a high-density built environment. The negative coefficient sign of the plot ratio suggests a lower value for housing units in denser developments. The positive sign of the external share implies the need for common facilities, ranging from the basics such as a stairway and an elevator to the luxuries and splendors like a mini golf course and a grand lobby. Certainly, adding a squared term could improve the model fit. However, it was not done, as model specification is not the paper’s focus.

On the contrary, the results on the green features are rather puzzling, without quantile regression. The coefficient estimate of the greening rate is negative in both OLS and 2SLS and significant in the latter, counter-intuitively suggesting that a greater coverage of the site area by greenery would result in lower house prices. Secondly, the model estimates also suggest a negative value placed on proximity to a natural park. This prediction is not straightforward: In Chinese cities, parks are the major means to carry out outdoor activities and provide enjoyable green views, as private backyards are scares. Quantile analysis will resolve these questions.

The remaining columns of Table 4 present the results of quantile regressions at the deciles, and the remaining columns of Table 5 display the outcomes of 2SQRs. The numbers in the parentheses are the bootstrapping standard errors, which are obtained through 500 bootstrap replications. Both tables exhibit similar patterns of changes in coefficient values across the deciles, with pseudo $R^2$ all well above 0.6.

Figure 2 includes 10 panels to support the illustration of how coefficient estimates of 10 explanatory variables, including the spatially lagged house price but excluding the district variables, vary across the conditional distribution of house prices. Each of these panels plots an explanatory variable’s 2SQR coefficient estimates and their associated 95% confidence intervals at 44 quantile points from the 6th to

---

7The district variables are not included, as the effect on prices of moving a unit from one district to another will not be discussed; such a location change can substantially alter the contents in the spatial weight matrix and the possible permutations are infinite.
94th percentile. The 2SLS estimate and bounds of its confidence interval are also included for comparison.

Panel 1 of Figure 2 as well as Table 4 reveals substantial variation of spatial dependence across quantiles. The value of $\hat{\lambda}_q$ is as high as 0.27 in the lower part of the response distribution, suggesting considerable positive spatial autocorrelation of house prices in that range. Moving to the right hand side of the distribution, the value decreases gradually, and it is statistically no different from zero between the 60th and 80th percentiles. After that, it picks up and approaches 0.25 at the 94th percentile. The 2SQR is more influential at those quantiles where spatial autocorrelation is more pronounced. Comparing Tables 3 and 4, it can be noted that 2SQR considerably changes the coefficient estimates and improves the model fit at the lower and upper deciles, but 2SQR makes little difference from the 6th to 8th deciles.

The above result suggests that the values of higher- and lower-priced homes are more positively influenced by the prices of nearby properties than the values of medium-priced homes. Such a U-shape pattern of spatial dependence across quantiles warrants more research. The pattern might be universal to datasets or specific to model design. On the other hand, the U-shape pattern could be specific to Changsha and the sample period, owing to China’s first housing voucher program, the Monetary Subsidy for Economical and Suitable Housing, implemented by Changsha city government in August 2008. As a part of the city’s economic stimulus package in response to the global economic crisis, the massive housing voucher program issued generous subsidy—about 40% of value of a new entry-level property. However, the program’s regulations pushed voucher recipients to purchase medium-priced homes. There was pressure of costly relocation resulted from the one-off subsidy that must be returned if the recipient sold the subsidized unit. Nevertheless, if the recipients felt that buying a medium-priced home too much of a stretch, they could still purchase a lower-priced unit. Given the market downturn, the enormous number of recipients\(^8\), who were required to spend their vouchers within one year, had a stake in the housing market. Thus, developers had a strong incentive to offer or accept discounts to attract these voucher holders. The incentive would be particularly strong for those developers selling medium-priced units in high-priced neighborhoods and could reduce spatial dependence over that price range.

In relative terms, the effect of change in characteristic $k$ on the house price at quantile $q$ is not $\hat{\beta}_{q,k}$.

In a spatial-lag model, a characteristic change of unit $i$ affects not only $i$’s price, but also the prices of $i$’s neighboring units, which may further influence some other units. Eventually, all these influences reciprocally impact $i$’s price. In other words, the model system has feedback loops. Thus, the effect on

---

\(^8\)According to the city government, there were about ten to sixteen thousands of households who received a voucher during the one year sample period.
$i$’s price due to $i$’s change in characteristic $k$ is represented by the $i$-th diagonal element of the matrix $\hat{\beta}_{q,k} \left[ I - \hat{\beta}_q W \right]^{-1}$. As different units are associated with different degrees and arrangements of spatial proximity to other units, the values of these diagonal elements are not identical. To construct useful discussion, the focus is being put on the average direct impact (LaSage and Pace, 2009), which is the average of the diagonal elements of the matrix.

Computationally, it is too heavy to invert a $46356 \times 46356$ matrix. Thus, we use the formula:

$$\left[ I - \hat{\beta}_q W \right]^{-1} \approx I + \sum_{m=1}^{M} \hat{\beta}_q^m W^m$$

where $M$ denotes the order of approximation. It is found that the second-order approximation performs sufficiently well in this case, as compared to higher-order approximations. In fact, LeSage and Pace (2009) show that the extra level of precision shrinks very quickly for the diagonal elements when moved to higher orders. Since every $\hat{\beta}_{q,k}$ is no more than 0.3% different from the value of the corresponding second order approximation of the average direct impact, we simply use it to interpret the percentage price change.

Panel 2 of Figure 2 depicts a clear upward trend of quantile effects of floor area, which is consistent with previous research. As the trend clearly crosses the confidence interval of the 2SLS estimate, using this estimate alone could overstate the price effects of floor area on lower-priced units but understate the effects on higher-priced units. In Panel 3, 2SQR estimates suggest that more bedrooms significantly increase house prices throughout the distribution. The coefficient is the largest at the first decile, which may be because households in low-priced homes tend to have a larger number of persons per bedroom. The coefficient peaks again around the 8th decile. A possible reconciliation is that high-income households, who are sorted into high-priced homes, have strong demand for a study room, entertainment room, playroom, etc.

Several covariates are particularly relevant to markets with condominiums as the major form of housing. Panel 4 of Figure 2 concerns the floor number. The 2SQR coefficient estimates are all significantly positive and depict a clear downward trend crossing the confidence interval of the 2SLS estimate. Panel 5 deals with the external share, the proportion of the owner’s property used as shared areas for common facilities. The coefficient estimates are all positive, and the pattern suggests particular importance of shared areas for lower-priced units. The estimates are quite constant throughout the medium price range, indicating the need for common areas. However, it decreases in the high price range, probably because wealthy households, who buy high-priced properties, value privacy more.
Panel 6 illustrates the results on the plot ratio. Other than the lower tail, the point estimates form a clear downward trend that crosses zero. Lower-priced units in taller buildings are generally associated with a premium, but higher-priced units bear a discount. The pattern cautions against using the 2SLS estimate to reflect an effect on prices for all units; the estimate suggests a negative expectation, but the effects are, in fact, positive at lower quantiles.

Intriguing results on the greening rate are displayed in Panel 7. Below the 50th percentile, the 2SQR estimates are positive. Beyond that, they are negative. A possible reason could be as follows. Chinese governments, including Changsha, often impose regulations against income segregation, thus, residential developments have a good mix of high- and low-priced units. If the public perceive that buyers of more expensive homes must bear more costs of provision of green space while all buyers equally benefit from it, then households who purchase lower-priced units would place a premium on a project with a higher greening rate but those who acquire higher-priced units would discount that. This can result in the pattern shown in Panel 7.

On the impact of a 1 km increase in the distance to the CBD, Panel 8 depicts a U-shape pattern. The intuition of the initial downward trend is straightforward. Higher-income households have higher opportunity cost of time. The reason for the later upward trend could be related to transportation modes. In Changsha, only the wealthy families have cars to benefit from the city’s convenient highway and expressway system. The majority of households rely on bus services with slow speed and insufficient route coverage outside the city’s central areas. Henceforth, the 2SQR coefficient estimates could bear an upward trend, along the upper part of the distribution where wealthier households increasingly switch from public transport to private vehicles. If using automobiles allows the rich to suburbanize but taking public transport constrains the poor in the central city, income segregation could become severe in the future (see Glaeser, Kahn and Rappaport, 2008). In this regard, the city’s current effort in constructing mass rapid transit seems a good initiative.

Panels 9 and 10 focus on the distance to urban and natural parks, respectively. Both panels portray clear downward trends, which are in line with the findings of previous research that demand for quality of living increases with income. Panel 9 shows that, a one kilometer closer to an urban park would command a roughly 4% premium on the price at the 90th percentile of the response distribution, but the effect is insignificant around the 10th percentile. Panel 10 shows that, in relative terms, the premium due to a 1 km shorter distance to a natural park is comparable to that of an urban park. However, below the

---

9 Examples are the “price-limited housing,” “90-70 rule,” and “association of public housing in commercial condominium.”

10 According to the 2009 Changsha Statistical Year Book, the number of private vehicles to households was 0.15 for the whole city. However, this ratio would be much higher in the city’s urbanized areas.

70th percentile, a shorter distance would result in a discount. A possible reason is that natural parks are usually in the outskirts, so lower-income households’ reliance on public transport could be a deterring factor.

Results in Panels 9 and 10 may also have policy implications. The city’s issues in park provision are the insufficiency and unevenness\(^\text{12}\). Its ratio of urban green spaces to land areas was once 15%, far below the central government’s target. Its green spaces are heavily concentrated in only a few large parks. For instance, the 135 hectares of land of Lieshi Park in Furong district make up 67% of the total green spaces in the city’s central area. Such heavy concentration, which is due to the city’s previous strategy of park provision, leads to unequal accessibility to green spaces across neighborhoods. The city has reformed its park provision strategy. More attention is put in constructing small community parks throughout the urban areas, such that every housing unit will be located within about 500 meters of a park, eventually. As the value of park proximity is expected to rise with city’s economic development and household income, the policy shift to put more emphasis on small community parks can insert much positive effect on prices into the housing market. Also, the policy shift could prevent income segregation.

5 Conclusion

Using a new dataset on an emerging Chinese city—Changsha, this paper applies spatial quantile regression to examine how implicit prices of housing characteristics vary across the conditional distribution of house prices as well as to control the effects of spatial dependence. With a large sample in hand, a two stage method, 2SQR, is applied to estimate a spatial-lag model at various quantiles. The estimates show that implicit prices of certain housing characteristics can vary considerably across the distribution, suggesting the usefulness of estimating the conditional quantile functions in addition to the conditional mean. Particularly, the value of the distance to the CBD exhibits a U-shape pattern along the quantiles. This may be related to households’ switch from public transport to private vehicles. The value of the distance to the nearest park has a downward trend, which has implications to park provision. Future applications of quantile regression to hedonic models can facilitate a better understanding of house price valuation and planning. Quantile regression may also help with consideration of missing variables.

In contrast to the past literature, this research finds the integration of spatial econometrics and quantile regression helpful, because the estimated spatial dependence varies substantially across quantiles. For Changsha, the degree of dependence is considerable in the upper and lower parts of the house price

\(^{12}\)See Changsha’s Planning Bulletin of Urban Green Space System.
distribution and thus has strong impacts on the point estimators for those ranges. Spatial dependence, however, is insignificant in the medium range. The identified U-shape pattern along the response distribution warrants future research to examine its robustness across datasets and models. Also, future works can identify better model specifications and provide rigorous statistical references for spatial quantile regression.

References


Figure 1: Changsha district map and the locations of the developments, parks and CBD.
Figure 2: Quantile effects, 2SQR

Panel 1: Spatially lagged house price

Panel 2: Floor area

Panel 3: Bedroom

Panel 4: Floor number

Panel 5: External share

Panel 6: Plot ratio

Panel 7: Greening rate

Panel 8: CBD distance

Panel 9: Urban park distance

Panel 10: Natural park distance
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>House price</td>
<td>Price of the trasacted housing unit</td>
</tr>
<tr>
<td>Floor area</td>
<td>The unit's total construction area including private living area and shared common area, in square meters</td>
</tr>
<tr>
<td>Bedroom</td>
<td>Number of bedrooms in the unit</td>
</tr>
<tr>
<td>Floor number</td>
<td>The building's storey on which the unit is located.</td>
</tr>
<tr>
<td>External share</td>
<td>Percentage share of the unit's shared common area in its total construction area</td>
</tr>
<tr>
<td>Plot ratio</td>
<td>The ratio of total floor area of the entire development to its site area</td>
</tr>
<tr>
<td>Greening rate</td>
<td>Percentage share of the site area that is covered by greenery</td>
</tr>
<tr>
<td>CBD distance</td>
<td>Distance to Chansha CBD, in kilometers</td>
</tr>
<tr>
<td>Urban park distance</td>
<td>Distance to the nearest park, in kilometers, if it is an urban park; 0 otherwise</td>
</tr>
<tr>
<td>Natural park distance</td>
<td>Distance to the nearest park, in kilometers, if it is an natural park; 0 otherwise</td>
</tr>
<tr>
<td>Yuhua</td>
<td>1 if the unit is in Yuhua district; 0 otherwise</td>
</tr>
<tr>
<td>Furong</td>
<td>1 if the unit is in Furong district; 0 otherwise</td>
</tr>
<tr>
<td>Kaifu</td>
<td>1 if the unit is in Kaifu district; 0 otherwise</td>
</tr>
<tr>
<td>Yuelu</td>
<td>1 if the unit is in Yuelu district; 0 otherwise</td>
</tr>
</tbody>
</table>
Table 2: Summary statistics of individual variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>House price</td>
<td>426738</td>
<td>214058</td>
<td>57056</td>
<td>3471885</td>
</tr>
<tr>
<td>Floor area</td>
<td>95.63</td>
<td>39.32</td>
<td>17.83</td>
<td>380.81</td>
</tr>
<tr>
<td>Bedroom</td>
<td>2.19</td>
<td>0.95</td>
<td>1.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Floor number</td>
<td>13.08</td>
<td>7.95</td>
<td>1.00</td>
<td>35.00</td>
</tr>
<tr>
<td>External share</td>
<td>20.02</td>
<td>4.18</td>
<td>6.72</td>
<td>35.62</td>
</tr>
<tr>
<td>Plot ratio</td>
<td>3.97</td>
<td>1.76</td>
<td>1.40</td>
<td>8.99</td>
</tr>
<tr>
<td>Greening rate</td>
<td>40.72</td>
<td>9.12</td>
<td>15.00</td>
<td>64.00</td>
</tr>
<tr>
<td>CBD Distance</td>
<td>6.27</td>
<td>2.45</td>
<td>0.64</td>
<td>13.91</td>
</tr>
<tr>
<td>Urban Park Distance</td>
<td>1.76</td>
<td>1.10</td>
<td>0.00</td>
<td>4.47</td>
</tr>
<tr>
<td>Natural Park Distance</td>
<td>0.23</td>
<td>1.05</td>
<td>0.00</td>
<td>6.17</td>
</tr>
<tr>
<td>Yuhua</td>
<td>0.32</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Furong</td>
<td>0.11</td>
<td>0.31</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Kaifu</td>
<td>0.22</td>
<td>0.41</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Yuelu</td>
<td>0.16</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3: OLS and QR estimates

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0104</td>
<td>0.0102</td>
<td>0.0104</td>
<td>0.0104</td>
<td>0.0106</td>
<td>0.0107</td>
<td>0.0108</td>
<td>0.0107</td>
<td>0.0105</td>
<td>0.0113</td>
</tr>
<tr>
<td>Floor area</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Bedroom</td>
<td>0.0189</td>
<td>0.0333</td>
<td>0.0102</td>
<td>0.0085</td>
<td>0.0012</td>
<td>0.0006</td>
<td>-0.036</td>
<td>0.063</td>
<td>0.0249</td>
<td>0.0093</td>
</tr>
<tr>
<td>Floor number</td>
<td>0.0029</td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0034</td>
<td>0.0032</td>
<td>0.0029</td>
<td>0.0029</td>
<td>0.0026</td>
<td>0.0022</td>
<td>0.0019</td>
</tr>
<tr>
<td>External share</td>
<td>0.0035</td>
<td>0.0048</td>
<td>0.0020</td>
<td>0.0036</td>
<td>0.0028</td>
<td>0.0033</td>
<td>0.0029</td>
<td>0.0035</td>
<td>0.0027</td>
<td>0.0020</td>
</tr>
<tr>
<td>Plot ratio</td>
<td>-0.0033</td>
<td>0.0063</td>
<td>0.0088</td>
<td>0.0022</td>
<td>0.0009</td>
<td>-0.0015</td>
<td>-0.0029</td>
<td>-0.0075</td>
<td>-0.0107</td>
<td>-0.0182</td>
</tr>
<tr>
<td>Greening rate</td>
<td>-0.0004</td>
<td>0.0019</td>
<td>0.0024</td>
<td>0.0014</td>
<td>0.0010</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0011</td>
<td>0.0024</td>
<td>0.0048</td>
</tr>
<tr>
<td>CBD distance</td>
<td>-0.0369</td>
<td>-0.0307</td>
<td>-0.0358</td>
<td>-0.0361</td>
<td>-0.0366</td>
<td>-0.0369</td>
<td>-0.0393</td>
<td>-0.0377</td>
<td>-0.0365</td>
<td>-0.0385</td>
</tr>
<tr>
<td>Urban park</td>
<td>-0.0214</td>
<td>-0.0118</td>
<td>-0.0069</td>
<td>-0.0156</td>
<td>-0.0208</td>
<td>-0.0230</td>
<td>-0.0206</td>
<td>-0.0252</td>
<td>-0.0286</td>
<td>-0.0410</td>
</tr>
<tr>
<td>Natural park</td>
<td>0.0080</td>
<td>0.0358</td>
<td>0.0304</td>
<td>0.0198</td>
<td>0.0144</td>
<td>0.0111</td>
<td>0.0106</td>
<td>0.0009</td>
<td>0.0131</td>
<td>0.0299</td>
</tr>
<tr>
<td>Yuhua</td>
<td>-0.0391</td>
<td>-0.0067</td>
<td>-0.0083</td>
<td>-0.0114</td>
<td>-0.0282</td>
<td>-0.0440</td>
<td>-0.0703</td>
<td>-0.0759</td>
<td>-0.0633</td>
<td>-0.0736</td>
</tr>
<tr>
<td>Furong</td>
<td>-0.0523</td>
<td>-0.0397</td>
<td>-0.0380</td>
<td>-0.0482</td>
<td>-0.0385</td>
<td>-0.0472</td>
<td>-0.0685</td>
<td>-0.0757</td>
<td>-0.0429</td>
<td>-0.0704</td>
</tr>
<tr>
<td>Kaifu</td>
<td>-0.0851</td>
<td>-0.0412</td>
<td>-0.0706</td>
<td>-0.0785</td>
<td>-0.0915</td>
<td>-0.0973</td>
<td>-0.1126</td>
<td>-0.1251</td>
<td>-0.1273</td>
<td>-0.1298</td>
</tr>
<tr>
<td>Yuelu</td>
<td>-0.1422</td>
<td>-0.1322</td>
<td>-0.1166</td>
<td>-0.1151</td>
<td>-0.1239</td>
<td>-0.1453</td>
<td>-0.1781</td>
<td>-0.1783</td>
<td>-0.1649</td>
<td>-0.1540</td>
</tr>
<tr>
<td>Adjusted/ Pseudo R²</td>
<td>0.8571</td>
<td>0.6135</td>
<td>0.6371</td>
<td>0.6383</td>
<td>0.6389</td>
<td>0.6394</td>
<td>0.6386</td>
<td>0.6333</td>
<td>0.6267</td>
<td>0.6308</td>
</tr>
</tbody>
</table>

Note: The dependent variable is the log-transformed house prices.
The sample size is 46356
** indicates 5% level of significance
In the parentheses under QR estimates are bootstrapping standard errors obtained through 500 bootstrap replications
Table 4: 2SLS and 2SQR estimates of the spatial lag model

<table>
<thead>
<tr>
<th></th>
<th>2SLS</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatially lagged house price</td>
<td>0.0986 ** (0.0074)</td>
<td>0.2593 ** (0.0254)</td>
<td>0.1637 ** (0.0104)</td>
<td>0.1652 ** (0.0111)</td>
<td>0.1088 ** (0.0144)</td>
<td>0.0536 ** (0.0117)</td>
<td>0.0048 (0.0106)</td>
<td>0.0141 (0.0112)</td>
<td>0.1407 ** (0.0170)</td>
<td></td>
</tr>
<tr>
<td>Floor area</td>
<td>0.0103 ** (0.0000)</td>
<td>0.0100 ** (0.0001)</td>
<td>0.0102 ** (0.0001)</td>
<td>0.0104 ** (0.0001)</td>
<td>0.0106 ** (0.0001)</td>
<td>0.0107 ** (0.0001)</td>
<td>0.0105 ** (0.0001)</td>
<td>0.0111 ** (0.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedroom</td>
<td>0.0184 ** (0.0020)</td>
<td>0.0299 ** (0.0060)</td>
<td>0.0071 ** (0.0035)</td>
<td>0.0106 ** (0.0032)</td>
<td>0.0051 (0.0025)</td>
<td>0.0033 (0.0028)</td>
<td>0.0064 (0.0035)</td>
<td>0.0242 ** (0.0045)</td>
<td>0.0065 (0.0051)</td>
<td></td>
</tr>
<tr>
<td>Floor number</td>
<td>0.0028 ** (0.0001)</td>
<td>0.0035 ** (0.0002)</td>
<td>0.0035 ** (0.0001)</td>
<td>0.0032 ** (0.0001)</td>
<td>0.0029 ** (0.0001)</td>
<td>0.0029 ** (0.0001)</td>
<td>0.0026 ** (0.0001)</td>
<td>0.0022 ** (0.0002)</td>
<td>0.0018 ** (0.0002)</td>
<td></td>
</tr>
<tr>
<td>External share</td>
<td>0.0038 ** (0.0003)</td>
<td>0.0064 ** (0.0005)</td>
<td>0.0026 ** (0.0004)</td>
<td>0.0042 ** (0.0003)</td>
<td>0.0035 ** (0.0003)</td>
<td>0.0034 ** (0.0003)</td>
<td>0.0029 ** (0.0003)</td>
<td>0.0027 ** (0.0004)</td>
<td>0.0012 ** (0.0006)</td>
<td></td>
</tr>
<tr>
<td>Plot ratio</td>
<td>-0.0041 ** (0.0007)</td>
<td>0.0081 ** (0.0018)</td>
<td>0.0044 ** (0.0012)</td>
<td>0.0013 ** (0.0003)</td>
<td>-0.0004 ** (0.0009)</td>
<td>-0.0024 ** (0.0008)</td>
<td>-0.0075 ** (0.0012)</td>
<td>-0.0111 ** (0.0013)</td>
<td>-0.0209 ** (0.0016)</td>
<td></td>
</tr>
<tr>
<td>Greening rate</td>
<td>-0.0007 ** (0.0001)</td>
<td>0.0011 ** (0.0003)</td>
<td>0.0016 ** (0.0002)</td>
<td>0.0010 ** (0.0002)</td>
<td>0.0007 ** (0.0002)</td>
<td>0.0000 ** (0.0002)</td>
<td>0.0002 ** (0.0002)</td>
<td>-0.0011 ** (0.0003)</td>
<td>-0.0061 ** (0.0003)</td>
<td></td>
</tr>
<tr>
<td>CBD distance</td>
<td>-0.0351 ** (0.0005)</td>
<td>-0.0232 ** (0.0016)</td>
<td>-0.0320 ** (0.0008)</td>
<td>-0.0337 ** (0.0008)</td>
<td>-0.0351 ** (0.0008)</td>
<td>-0.0364 ** (0.0007)</td>
<td>-0.0393 ** (0.0006)</td>
<td>-0.0377 ** (0.0009)</td>
<td>-0.0362 ** (0.0009)</td>
<td></td>
</tr>
<tr>
<td>Urban park distance</td>
<td>-0.0219 ** (0.0010)</td>
<td>-0.0096 ** (0.0015)</td>
<td>-0.0071 ** (0.0015)</td>
<td>-0.0119 ** (0.0014)</td>
<td>-0.0200 ** (0.0013)</td>
<td>-0.0228 ** (0.0010)</td>
<td>-0.0251 ** (0.0011)</td>
<td>-0.0292 ** (0.0020)</td>
<td>-0.0429 ** (0.0018)</td>
<td></td>
</tr>
<tr>
<td>Natural park distance</td>
<td>0.0070 ** (0.0011)</td>
<td>0.0328 ** (0.0027)</td>
<td>0.0248 ** (0.0012)</td>
<td>0.0213 ** (0.0011)</td>
<td>0.0150 ** (0.0013)</td>
<td>0.0114 ** (0.0010)</td>
<td>0.0106 ** (0.0009)</td>
<td>-0.0008 ** (0.0015)</td>
<td>-0.0336 ** (0.0016)</td>
<td></td>
</tr>
<tr>
<td>Yuhua</td>
<td>-0.0449 ** (0.0025)</td>
<td>-0.0281 ** (0.0078)</td>
<td>-0.0250 ** (0.0036)</td>
<td>-0.0213 ** (0.0030)</td>
<td>-0.0329 ** (0.0039)</td>
<td>-0.0466 ** (0.0037)</td>
<td>-0.0704 ** (0.0033)</td>
<td>-0.0760 ** (0.0045)</td>
<td>-0.0790 ** (0.0059)</td>
<td></td>
</tr>
<tr>
<td>Furong</td>
<td>-0.0617 ** (0.0033)</td>
<td>-0.0659 ** (0.0087)</td>
<td>-0.0467 ** (0.0066)</td>
<td>-0.0536 ** (0.0045)</td>
<td>-0.0395 ** (0.0048)</td>
<td>-0.0492 ** (0.0046)</td>
<td>-0.0685 ** (0.0046)</td>
<td>-0.0758 ** (0.0046)</td>
<td>-0.0449 ** (0.0079)</td>
<td></td>
</tr>
<tr>
<td>Kaifu</td>
<td>-0.0977 ** (0.0029)</td>
<td>-0.0857 ** (0.0084)</td>
<td>-0.0948 ** (0.0046)</td>
<td>-0.1000 ** (0.0036)</td>
<td>-0.1057 ** (0.0047)</td>
<td>-0.1016 ** (0.0047)</td>
<td>-0.1130 ** (0.0038)</td>
<td>-0.1254 ** (0.0042)</td>
<td>-0.1296 ** (0.0057)</td>
<td></td>
</tr>
<tr>
<td>Yuelu</td>
<td>-0.1414 ** (0.0032)</td>
<td>-0.1090 ** (0.0047)</td>
<td>-0.1029 ** (0.0041)</td>
<td>-0.1225 ** (0.0053)</td>
<td>-0.1274 ** (0.0047)</td>
<td>-0.1475 ** (0.0047)</td>
<td>-0.1782 ** (0.0053)</td>
<td>-0.1786 ** (0.0055)</td>
<td>-0.1651 ** (0.0075)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>10.8187 ** (0.0943)</td>
<td>8.2721 ** (0.3306)</td>
<td>9.7403 ** (0.1272)</td>
<td>9.7766 ** (0.1402)</td>
<td>10.5992 ** (0.1843)</td>
<td>11.3872 ** (0.1495)</td>
<td>12.0870 ** (0.1349)</td>
<td>12.2168 ** (0.1487)</td>
<td>12.1871 ** (0.1364)</td>
<td>10.8592 ** (0.2085)</td>
</tr>
</tbody>
</table>

Note: The dependent variable is the log-transformed house prices.
The sample size is 46356
** indicates 5% level of significance
In the parentheses under 2SQR estimates are bootstrapping standard errors obtained through 500 bootstrap replications