UNDERSTANDING CHANGES IN THE DISTRIBUTION OF HOUSE PRICES IN BEIJING, CHINA

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Yimin Yi

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Agribusiness and Applied Economics

June 2016

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

UNDERSTANDING CHANGES IN THE DISTRIBUTION OF HOUSE PRICES IN BEIJING, CHINA

By

Yimin Yi

The supervisory committee certifies that this thesis complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Prof. Lei Zhang

Chair

Prof. Saleem Shaik

Prof. Jun Kong

Approved:

22 June 2016

Date

Prof. William Nganje

Department Chair

ABSTRACT

This thesis provides some evidences in understanding house price in Beijing. The first part examines appreciation rate of house price across the distribution from 2013 to 2015. Similar temporal patterns of appreciation for different parts of price distribution are shown, while the rates of appreciation of low-priced homes are found to be higher than higher-priced homes over almost the full research period.

The second part analyzes changes in the distribution of house price between 2012 and 2015. We disentangle temporal changes into a composition effect attributed to altered house characteristics and a coefficient effect driven by varying regression coefficients. Mean decomposition suggests that only 13% of average price gaps are attributed to the composition effect. Quantile decomposition results indicate that the contribution of the composition effect rises monotonously from the left tail of distribution to the right tail while the contribution of the coefficient effect only shows slight variation.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my advisor, Dr. Lei Zhang, for her insight and patience. Her guidance given to me is the major driving force in the success of my study and this thesis. Hopefully, we can collaborate on a long time in the future in the area of Urban Economics. I also want to thank my committee members. Dr.Saleem Shaik and Dr.Jun Kong, for giving me their valuable suggestions to make this thesis completed.

Special thanks to my parents and relatives for their unremitting encouragement and supports in my life. My instructors and classmates here at NDSU also gave me excellent advices and supports on completing my study, many friends in Fargo provided me great help and courage to get through obstacles in life, thank you all. I also want to thank my roommate, Liming Xie, for the delicious food he cooked for me and the cooking skills I learned from him. Thanks also go out to mates in the soccer ground, a lot of names not listed there, for the enjoyable time I had been with them.

DEDICATION

This thesis is dedicated to my little brother, Chao Yi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

1. INTRODUCTION AND BACKGROUND

1.1. Problem Statement

According to World Bank, in 2014, China's gross domestic product reached unprecedented 18,017,073 million international dollars, contributed to roughly 16.5 percent of the global GDP. In the same time period, the GDP of the United States was 17,419,000 million international dollars. Therefore, China surpassed USA, ranked No.1 in all the countries around the world.[1] During the period from 1990 to 2010, the yearly growth rate of China's GDP reached about unprecedented 10.4 percent(Lin, 2011). As a result of improvement of status in the world economy, China has been a core role in the recovery of the world economy from the global financial crisis since 2008. But, recently, the growth rate of the Chinese economy seems to slow down, as the official annual growth rate of GDP announced to be less than 7 percent (6.9 percent) in 2015, which is the first time that since 1990[2]. Seemly return-to-normal growth has led to the global economic communities growing doubt and concerns that the Chinese economy may be a hard landing.

Accompanied by the rapid growth of Chinese economy, the housing market of China also has experienced a "golden" period in recent decades. A comprehensive reform of Chinese housing system was introduced in the early 1990s, resulting in change of housing allocation in China from employment non-market stepped into a private market of real estate. After that, the private housing market in China enjoyed dramatic growth since late 1990s. There has been a growing role of importance since the continuous housing market booming. As Deng et al. (2011) estimate, the real estate sector directly contributed roughly 11 percent of China's total GDP in 2009. The contribution rate may be even larger if related industries such as construction material industries were taken into consideration. In addition, land sales

[1]Source: http://databank.worldbank.org/data/download/GDP_PPP.pdf, World Development Indicators database, World Bank, 11 April 2016. GDP are calculated using purchasing power parity. Rankings based on only countries with confirmed PPP GDP estimates.

[2]http://www.tradingeconomics.com/articles/01192016032545.htm

revenue is an important source of local government revenue in China. In 2009, land sales revenue contributed almost 35 percent of local government revenues on average in China, the share even rose up to more than 45 percent in some places such as Tianjin and Chongqing (Lu and Sun, 2013). Rising price of real estate may accumulate high risk of housing bubble, also high dependence of government revenue on land sales revenue may cause the Chinese economy be bound by the housing industry. Wu et al. (2012) find that the property price-to-rent ratio has increased significantly in 8 Chinese cities since 2007. Thus, they suggest that there may be substantial mispricing in Chinese property market and high risk that large price depreciation may be caused by even modest fall in expected appreciation.

An accurate house price measurement is of great importance to detect potential real estate mis-pricing and even housing bubble, measure housing affordability, and assess the latent risk of economic crisis, not only for market participants, research community, but also for policy makers. As the capital and biggest city in China, house price in Beijing has captured more attention than any other cities and may serve as a representative of first-tier cities in China. A wide arrange of literature has paid attention on house price index issues, however, most of the literature has focused on mean house price indices, few attempt has been made to take the full distribution into consideration, i.e., to develop a measure of house price varies with locations in the distribution of price. So, firstly, we want to fill this gap to see whether there are heterogeneities on the appreciation rate of house price across the distribution in Beijing? Secondly, as we observed substantial increase in house price between 2012 to 2015 in Beijing, one natural question come into mind is what caused house prices to grow so quickly? Were increase in house price driven by simply more high-priced houses sold in 2015 or the upgrading of housing characteristics over time, or higher returns to certain characteristics from 2012 to 2015?

1.2. Objectives

This analysis is aimed to present some evidences in understanding house price in Bejing over recent years. Specifically, we want to know whether low-priced houses shared the same appreciation rate for middle-priced and high-price homes? And what caused the variation in the distribution of house price over time in Beijing? Main questions of this study are listed as follows:

- What are the determinants of house price in Beijing?
- Are there explanation power of the investigated characteristics in explaining house price?
- What is the pattern of evolution of house price over time?
- Did the temporal pattern of house price indices vary with positions in the distribution of house price?
- Did return to house characteristics change between 2012 and 2015?
- What are differences in the distribution of house prices between 2012 and 2015?
- How did changes in house characteristics contribute to gaps in house price over time?
- Did the composition effect change with the location of house price?

The theoretical foundation is the hedonic price theory and the empirical methods applied in this study are OLS regression, quantile regression, mean decomposition, and quantile decomposition. Hopefully, in answering these questions, our analysis can contribute to promote understanding of house price dynamics to at least the research communities, participants in the housing market, policy makers, and people who may be interested in house price of China.

1.3. The Housing Market in Beijing, China

Beijing, as the capital city of China, has become one of the largest global city in the world. Currently, Beijing city comprises 16 districts, of which 6 urban districts make up of the city proper. Dongcheng district and Xicheng district both are within the second ring-road, and they form the traditional core area of the city. The other four inner-suburban districts are Chaoyang, Fengtai, Shijingshan, and Haidian. The six districts together constitute an area of 1377.82 square kilometers and are called the "Chenliu Qu"(City of six districts), which is the city's central urban area. The left 10 districts are outer-suburban districts. These districts are in the area that the basic industry is moving outside from inner-urban districts into, as a result of growing urbanization in Beijing. The districts of the city of Beijing are shown in figure 1.1.

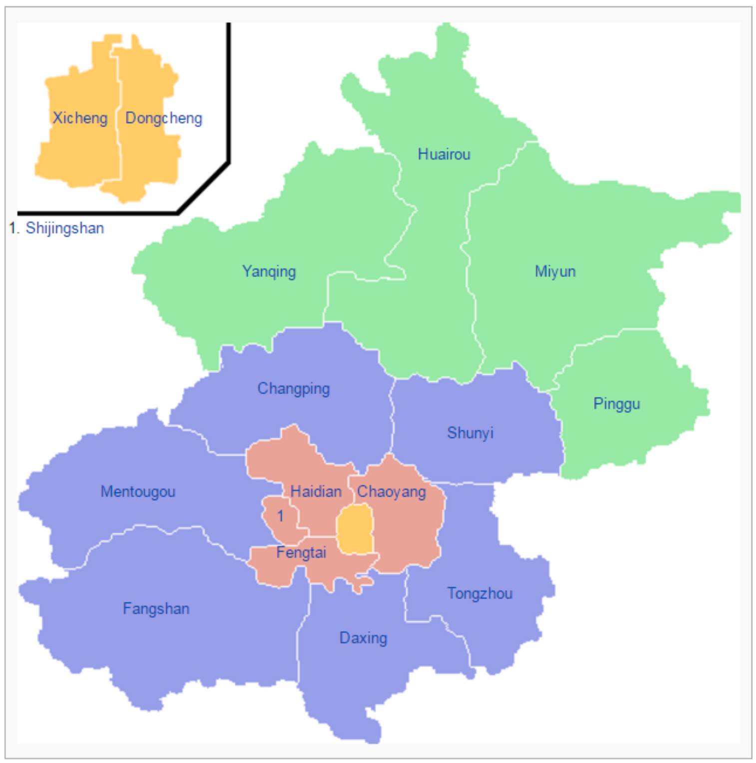

Figure 1.1. 16 Urban Districts of Beijing in 2016. Source: Wikipedia.

The permanent population of the city has increased from 10.86 million to 21.71 million from 1990 to 2014, nearly doubled in 24 years. The statistics show that 12.28 million of permanent populations live between the third ring-road and the sixth ring-road, while 10.98 million are outside the fifth ring-road, make up for 57.1 percent and 51 percent of the total permanent population, respectively. The implication is a growing trend that people have a tendency to live on the periphery of the city other than the inner city in consideration of living cost. The annual disposable income per capita (in constant RMB) has almost tripled from RMB 17.7 thousand in 2005 to RMB 48.5 thousand ($7,456 in US dollars) in 2015[3]. The rising number of the permanent population and disposable income led into a growing demand for housing service in Beijing.

The private housing market in Beijing dates back to early 1990s. Before that housing service was supplied by work units, who built and allocated housing units to their employees. Employees got housing units at low subsidized rent. In 1994, the Chinese government initiated a comprehensive housing reform. The reform built up a new housing provision system. In this system, housing units became commodities under the construction of development companies. Workers no longer get allocated housing service from their employers but to purchase houses from development companies. The subsidy of employee paid by a work unit for housing service adds up to workers' total salary. In Beijing, the majority of commodity housing usually not take the form of single houses but a complex-where many apartment-like housing units aggregate in the same medium-rise or high-rise building. Housing units in the same complex share almost the same environmental characteristics and geographic attributes. However, different complex may exhibit tremendous variation on location and location-related attributes, such as subway proximity, landscaping, and even building design.

After the reform, the private housing market in Beijing started to take off and an almost non-terminated booming market period came. During the period from 2002 to 2013, the annual growth in residential real estate sales maintained about 15 percent on average, and the annual growth of residential housing areas is even greater, reaching about 18 percent on average. In 2015, the areas of commodity residential housing units under construction reached up to 130.95 million square meters, while the corresponding areas in 1994 were only 11.076 million square meters.

[3] The statistics come from Beijing Municipal Bureau of Statistics: http://www.bjstats.gov.cn/zt/rkjd/.

1.4. Organization of Study

The remaining paper proceeds as follows. Chapter 2 presents a brief introduction of hedonic theory. We discuss heterogeneities on the appreciation rate of house prices across the distribution in Chapter 3. Following that we provide evidences for understanding what caused changes in the distribution of house prices over time in chapter 4. Chapter 5 concludes.

2. THEORETICAL FOUNDATION

In this chapter, we discuss the hedonic pricing theory. Section 2.1 dates back the history of hedonic models. Section 2.2 briefly introduces the equilibrium in a demand-supply hedonic model.

2.1. The History of Hedonic Models

It is not clear when talking about the “father” of the hedonic pricing model. A study by Goodman (1998) re-examines the work of Court (1939), which develops an automobile price index, points out that the term “hedonic” may date back to the study of Court (1939). Goodman (1998) argues that, even though the hedonic method popularized after Griliches (1958) on estimating the demand for fertilizer, the work of Court (1939) received more citations and stands up well even under contemporary standards. As an economist for the Automobile Manufacturers’Association in Detroit, Court (1939) thinks that one single variable may not explain automobile demand well, he estimates the demand of automobile from 1925 to 1939 using dry weight, wheelbase, and horsepower. But a study by Colwell and Dilmore (1999) raise question about who first employed hedonic approach, they suggest that a monograph by Haas (1922) employed a hedonic method to estimate the value of farmland even the term “hedonic” was put up by Court (1939). Also, four years later after Haas (1922), Wallace (1926) carried a similar study to Haas (1922), he appraised farmland value in Iowa by using a hedonic technique. However, the theoretical foundation of the hedonic price model was introduced by Lancaster (1966), which provides a microeconomic basis for estimating the implicit price of utility-bearing characteristics. Rosen (1974) advances the theory by focusing on the characteristics that determine the price of a commodity while not only emphasis on the generation of utility. Also, Rosen’s work provides the basis for non-linear hedonic pricing estimation.

2.2. Hedonic Theory

Following the work of Lancaster (1966) and Rosen (1974), consider a generic commodity that has a bunch of attributes Z, where $Z = (z_1, z_2, \ldots z_n)$ are varying quantities of attributes that generate utility for the consumer. Rosen (1974) assumes for each consumer, there exists a bid function θ which defines the maximum willingness to pay by these consumers for the commodity containing bundle Z. In addition, the consumer's bid function θ, given a bundle of attributes, depends upon the consumer's income level I and preferences. That is,

$$\theta = \theta(Z, I, \alpha) \tag{2.1}$$

where α is a vector of characteristics that determine the preference of the consumer. Further assuming that the utility function is concave and all attributes of the generic commodity are normal goods, then the customer's bid or value function is an increasing function of any commodity attributes at a decreasing rate.

The first order partial derivative of the consumer bid function with respect to any commodity characteristic z_i, $\partial\theta/\partial z_i$, indicates the consumer's implicit bid for the commodity characteristic z_i. Under the assumptions above, increase on the quantities of the attributes would result in decreasing on the implicit bid.

On the supply side, symmetrically, Rosen defines an offer function, ϕ, which is defined as the minimum unit payment the producer is willing to accept the commodity it produces. Assuming producers are rationally maximize profit seeker and the market is competitive, the producer's offer function, ϕ, is affected by the attributes of commodity $Z = (z_1, z_2, \ldots z_n)$, the production level M, prices of factor, and production technology. Formally,

$$\phi = \phi(Z, M, \beta) \tag{2.2}$$

where β is a vector of factor prices and production parameters. If we further assume the producer's profit function is convex, then the offer function ϕ would be expected to be constant or increasing as the quantity of any commodity characteristics increase, i.e., if the amount of any commodity characteristic z_i increase, the implicit offer price for that characteristic $\partial\phi/\partial z_i$ would be constant or increasing.

Market equilibrium happens when the bid functions and offer functions for the attributes of the generic commodity are tangent, i.e.,

$$Q^s(Z) = Q^d(Z) \tag{2.3}$$

where the market the commodity bundle is clear. In addition, it is also necessary that any sub-market for the n commodity characteristics to be in equilibrium, i.e.,

$$Q^s(z_i) = Q^d(z_i),\ for\ i = 1, 2, \ldots, n. \tag{2.4}$$

or

$$\phi(z_i) = \theta(z_i),\ for\ i = 1, 2, \ldots, n. \tag{2.5}$$

Then we can have a system of $2n$ simultaneous equations derived from the n markets for the n commodity characteristics. The simultaneous solution to these $2n$ equations also guarantees that the market for the n commodity bundles is also in equilibrium.

Hedonic pricing theory suggests that consumers derive utility not directly from consumption of the actual good or service but from a bundle of characteristics embodied in that good or service. The implicit prices of the bundle of characteristics associated with the commodity are then defined as hedonic prices. For example, the price of a house can be connected to specific quantities of diverse characteristics such as number of bedrooms, number of garages, square footage etc.

3. QUANTILE HOUSE PRICE INDICES

3.1. Introduction

Currently, official house price indices are regularly reported by the National Bureau of Statistics of China (NBSC thereafter) in two forms at the same time: "Price indices for Real Estate in 35/70 Large and Medium Sized Cities" and "Average Selling Price of Newly Built Residential Buildings", and are commonly abbreviated as "NBS 70 Cities Index" and "NBS Average Price Index", respectively. The NBS 70 Cities Index is released monthly since July 2005. In each month, the average transaction price of housing units within the sample development is calculated and then compared with the average price for the same development in the preceding month to get the appreciation rate. The monthly house price appreciation rate at the city level is then calculated as the average, weighted by transaction volume, of the appreciation rate of all sampled developments in the corresponding month. The NBS Average Price Index is calculated by simply dividing the total sales price of all newly-built sold housing units by the total volume (in floor area) of the sold units for each sample month and given city. Though these two indices are both reported by NBSC, they are sometimes in conflict with each other. Both two official series have commonly known problems and have been wildly criticized for potential underestimate of actual real estate price appreciation, especially the NBS 70 Cities Index(see, e.g., Wu et al., 2014, 2015; Fang et al., 2015). As pointed out in Wu et al. (2015) and Fang et al. (2015), the NBS 70 Cities Index shows little variation of house price and moreover very little appreciation of real estate price in sampled 70 cities from 2000 to 2010. While the NBS Average Price Index is only an average transaction price, it does not take into consideration the fact that the house market in China is experiencing substantial changes. For example, developments built in later years are most likely to be farther away from the city center as a result of urbanization in most Chinese cities.

A growing body of literature has shed light on the issue of construction of house price indices, most of them whether through hedonic price model (see Hill, 2013, for a survey of hedonic price indices), or through repeat-sales method or its variations (see, e.g., Clapp and Giaccotto, 1992; Wang and Zorn, 1997; Harding et al., 2007; Deng et al., 2012). Wallace and Meese (1997) estimate the house price indices for Oakland and Fremont California from 1968 to 1990, suggesting that hedonic model provides better estimation than repeat-sales method since repeat-sales methods may suffer the problem of sample selection bias and be sensitive to extreme observations. Moreover, because the assumption of constant attributes of house unit over time cannot be satisfied, hedonic approach may be better suited than repeat-sales methods as the changing attributes over time are taken into account in hedonic techniques. The recent survey of Hill (2013) also concludes that hedonic indices seems to dominate current literature of house indices studies as data set that contains detailed attributes and characteristics of house units become available, since the problem of omitted variables is the main concern of hedonic housing studies. Most of the exist house indices studies focus on the secondary housing markets, only little attention has been paid to nascent markets like China where newly-built house units dominate the real estate market until recently (Guo et al., 2014; Wu et al., 2014; Fang et al., 2015). Based on the unique feature of the Chinese housing market, Wu et al. (2014) compare three common construction methods of house price index, namely the simple average method that does not accommodate housing quality, a matching approach, and the hedonic model. Their findings suggest that only the hedonic method considers both temporal quality change and developers' pricing strategy. Thus the hedonic method is the most suitable for the construction of house price indices in a nascent real estate market like China.

While the house price index constructed by either a hedonic approach or a repeat sales method does provide a quality-adjusted measure of house price, their interpretation sometimes causes some doubt(McMillen, 2008). For example, if the house price index rise from 1 in period 1 to 1.2 in period 2, the explanation is the price of a representative house increased by 20 percent(approximately). This implies that all houses in the market appreciate at the same level, or at least the appreciation rate are randomly distributed around the estimation of mean value. In reality, this may be not the truth. There may be substantial variation on the appreciation rate of house price across the full distribution for a given period or even great difference on the temporal pattern of house price dynamics for different percentages of house price. Only few works have paid attention to the heterogeneity of house appreciation rate in the past literature, to the best of our knowledge. Coulson and McMillen (2007) provide empirical evidence of constructing quantile real estate price indices for three suburban municipalities in the Chicago area. Their study identifies significant variations in the implicit price of physical characteristics across conditional quantiles and also disparity of appreciation rate among different quantiles.

In this chapter, we mainly concentrate on the temporal pattern of house price index across regression quantiles. Using a recent comprehensive micro-level data set of newly-built residential housing units in Beijing, this paper first examines the determinants of house price from 2013 to 2015. Our findings suggest that appreciation rate of house price is significantly higher from 2013 to 2015 in Beijing, which may imply even increasing risk of mispricing in the current housing market in China. Moreover, we find that most of housing attributes are valued differently by different percentage price of home buyers. Then we construct house price indices both from OLS and quantile regression estimation. The temporal pattern of house price index is similar across different quantile of the house price distributions, but various magnitude of appreciation rate across different percentiles is found among all the research period.

This research differs from most previous research on house price indices in that we take into account the full distribution of house prices. Our work contributes to the literature in two folds. Firstly, we obtained the data set of housing information from a leading Chinese housing website using data mining technique, which is not commonly applied in most of previous literature. As the information technology advance in developing countries like

China, where micro-level data are often not available or hard to access, data mining approach may equip us with a powerful tool to understand preference and decision making details of consumers. This may promote a direction of utilizing information on the Internet to better understand the sweep of economic trends especially in housing related area. In addition, this study provides the first quantile house price index in China. Our finding may not only provide valuable information to home buyers, research community, and policy makers to better understand the housing market conditions, but also provide convenience or direction for more thorough empirical analysis on housing issues.

The remaining chapter is organized as follows. We discuss the theoretical model in section 3.2. Following that is the estimation approach in section 3.3. Section 3.4 describes the procedure of data mining and presents summary statistics. Section 3.5 provides estimation results and the final section concludes.

3.2. Hedonic Pricing Model

The hedonic pricing model is widely applied in house price studies because it is a revealed preference method and the decomposition of housing value into price of individual characteristics is quite convenient. A typical hedonic model estimates implicit prices of attributes of house prices take the following form

$$p_{ijt} = \alpha + \boldsymbol{\beta}'\mathbf{U}_{it} + \boldsymbol{\gamma}'\mathbf{C}_{jt} + \boldsymbol{\phi}'\mathbf{G}_{jt} + \boldsymbol{\delta}'\mathbf{M}_{it} + \epsilon_{ijt} \tag{3.1}$$

where p_{ijt} is the transaction price of house i in complex j at time t, which typically take the natural logarithm form; $\mathbf{U}_{it}$ is a vector of unit-level attributes for house i, e.g. number of bedrooms, floor number, and area; $\mathbf{C}_{jt}$ is the complex-level characteristics such as greening rate and floor are ratio; $\mathbf{G}_{jt}$ and $\mathbf{M}_{it}$ are sets of geographic attributes and time dummies(equal 1 when the house is sold during the period t, 0 otherwise), respectively; ϵ_{ijt} is error term. $\boldsymbol{\delta}$ equals $\{\delta_2, \delta_3, \ldots, \delta_T\}'$ are the estimated price indexes for time period range from 2 to T.

Table 3.1. Variable with Inconsistent Estimated Coefficients across Studies.

Variables	Appearance	Positive times	Negative times	Insignificant times
Age	78	7	63	8
Distance	15	5	5	5
Bedrooms	40	21	9	10
Time on market	18	1	8	9

Some variables are expected to affect the house price positively. For example, as the area of house increase, the house price is going to be higher. While some other attributes may be detrimental to house price. The living quality is expected to be impaired as the floor area ratio of a development increase. So the house price falls when the floor area ratio rises. However, it is not uncommon to see the coefficient estimates differ in the magnitude, significance level and even sign in recent researches. Sirmans et al. (2005) summarizes approximately 125 recent studies applying hedonic pricing model and finds that some difference even exists for common characteristics. Predominantly inconsistent results are summarized in Table 3.1. For example, 63 of 78 studies report that the estimated coefficients of age are negative, while 7 of these studies show a positive effect. The remaining 8 other papers report insignificance results.

For explanations of the diverse results, the most obvious reason may be due to the difference of study places. People in the town may have different preference over some characteristics compared with people in the big city. Chinese may value some attributes of house differently from other nations. In addition, spatial dependence that observation at place A depend on other location B may affect the value of house price since its presence harms unbiasdness and efficiency of OLS estimates. The most possible reason may be that housing attributes are evaluated at different percentage of conditional house prices, which can partially explain the difference of hedonic price estimates in literatures. In this regard, quantile regression (Koenker and Bassett Jr, 1978) is quite suitable for house-related studies since it takes the whole conditional distribution of house prices into account rather than the mean only.

3.3. Quantile Regression

In classical OLS regression, conditional mean functions are estimated by minimizing sum of the square residuals,

$$\min_{\{\beta_i\}_{i=0}^{k}} \sum_j (y_j - \sum_{i=0}^{k} \beta_i x_{ij})^2 \tag{3.2}$$

where y_j is the jth observation of dependent variable, x_{ij} is the ith independent variable at observation j, β_i is the estimate of ith regression coefficient. While quantile regression take the entire distribution into account to estimate conditional quantile functions based on minimizing the sum of asymmetrically weighted absolute residuals (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001). In contrast, the minimization problem of a quantile regression can be written as:

$$\min_{\hat{\beta}_\rho} \sum_j w_j |y_j - \sum_{i=0}^{k} \beta_i x_{ij}| \tag{3.3}$$

where $\hat{\beta}_\rho$ is the vector of coefficient estimates, $\rho \in (0, 1)$ denotes the quantile to be estimated. The weight w_j is defined as:

$$w_j = 2\rho \tag{3.4}$$

if $y_j - \sum_{i=0}^{k} \beta_i x_{ij} > 0$, and

$$w_j = 2 - 2\rho \tag{3.5}$$

otherwise. The standard errors of coefficient estimates $\hat{\beta}_\rho$ can be calculated using bootstrapping. Only for the median ($\rho = 0.5$), symmetric weights are used, otherwise (e.g., $\rho = 0.1, 0.25 \ldots, 0.95$) asymmetric weights are used. Unlike OLS, quantile regression can be applied to estimate determinants of explained variable at any positions in the distribution of the dependent variable.

Although quantile regression has been widely used in empirical economic literature and provide persuasive explanations and instructive policies for the confusing society (see Koenker and Hallock, 2001), its application on house prices researches using hedonic price method extends until recently (Coulson and McMillen, 2007; McMillen, 2008; Zietz et al., 2008; Farmer and Lipscomb, 2010; Mak et al., 2010; Zhang and Leonard, 2014). McMillen (2008) construct quantile house price index to study the appreciation rates of house price in Chicago, find significantly variation of house characteristics value across conditional quantiles. Using quantile regression and incorporating spatial autocorrelation into hedonic house price model, Zietz et al. (2008) also show some housing characteristics such as square footage and the number of bathroom are valued differently between upper-quantile house and lower-quantile house.

3.4. Data

The dataset for empirical estimation was mined automatically by a python program from a public access Chinese housing website *Soufun*[1]. The website reports both descriptive information of complex characteristics and transaction records for newly-built houses. Descriptive information of complex and transaction record of the housing unit was matched for each observation. The sample contains 190580 observations of newly-built housing transactions for the period from 2013 to 2015 after data cleaning. Information of housing transaction is grouped into four categories: unit-level attributes, complex-level characteristics, geographic attributes, and monthly sold dummies. Table 3.2 lists variables available in the dataset according to each category. For each transaction, we have unit-level variables including transaction price, number of bedrooms, number of living rooms, size of living area, floor number, villa dummy, and face directions of the unit, the base group for the face direction is north; complex-level variables containing greening space ratio, floor area ratio, whether the complex has a sightseeing pool, and whether the complex a high-rise building; geographic attributes including whether the complex is near a subway station, whether the complex is

[1] The address of the website: http://fang.com/

near a park, and relative location to a ring road, the base group is housing units located outside of the sixth ring road; and monthly sold dummy that is determined by the transaction month.

The average of natural logarithm nominal sale price is 5.842, i.e. house price averaged 344.468 (10,000 Chinese Yuan) for the full period. The average number of bedrooms, number of living rooms, area, and floor number are 2.262, 1.441, 111.216, and 14.809, respectively. There are 6.7% of houses are villas. About half of the units (49.5%) face both south and north (southnorth), while the second most direction is south (16.2%). There are 3.6%, 4.7%, 2.7%, 3%, 2.8%, 5.7%, and 7.4% of housing units face west, east, north and west, north and east, east and west, south and west, and south and east, respectively. For the complex-level characteristics, the average greening rate is 0.349. Floor area ratio averages 2.515. Only 1.5 percent of complexes have a sightseeing pool inside while 73.6% of complexes are high-rise buildings. Regarding geographic attributes, we can see that 36.5% of complexes are near a subway station while only 4.5% are near a park. Most of the sold houses are outside of the third ring road, e.g. there are 27% of houses are located between the fifth ring road and the sixth ring road, which may be seen as a reflection of suburbanization. Month time dummies is the portion of total houses sold at this month, where the base time period is January 2013. For example, the mean value of Month 2 is 0.008, i.e., 8 percent of total housing units were sold in February 2013. The summary statistics of each variable are listed in table 3.3.

3.5. Estimation Results

3.5.1. OLS and Quantile Estimates

First, we estimate the baseline hedonic model via ordinary least square regression. The dependent variable is the natural logarithmic of the transaction price of each house(in 10,000 Chinese Yuan). The explanatory variables are divided into four categories, including unit-level attributes, complex-level characteristics, geographic attributes, and time dummies(month of the sale, start from January 2013, which is the base period).

Table 3.2. Definition of Variables.

Variable	Definition
1.Unit-level attributes	
lnprice	Transaction price(10,000 RMB Yuan), in logarithm form
Bedrooms	Number of bedrooms
DiningRooms	Number of dining rooms
area	Room area of the house, in square meters
floor	floor number
Villa	whether the unit a villa; 1 = Yes, otherwise 0
North	1 if the unit faces north, otherwise 0
West	1 if the unit faces west, otherwise 0
East	1 if the unit faces east, otherwise 0
South	1 if the unit faces south, otherwise 0
Northwest	1 if the unit faces north and west, otherwise 0
Northeast	1 if the unit faces north and east, otherwise 0
Eastwest	1 if the unit faces east and west, otherwise 0
Southwest	1 if the unit faces south and west, otherwise 0
Southeast	1 if the unit faces south and east, otherwise 0
Southnorth	1 if the unit faces south and north, otherwise 0
2.Complex-level characteristics	
Green	Greening space rate of the complex
FAR	Floor area ratio of the complex
WaterView	whether the complex has a sightseeing pool inside; 1 = Yes, otherwise 0
HighRise	whether the complex a high-rise building; 1 = Yes, otherwise 0
3.Geographic attributes	
NearSubway	whether the complex is near a subway station; 1 = Yes, otherwise 0
NearPark	whether the complex is near a park; 1 = Yes, otherwise 0
ringRoad02	1 if the complex is within the 2nd ring road; otherwise 0
ringRoad23	1 if the complex is between the 2nd and the 3rd ring road; otherwise 0
ringRoad34	1 if the complex is between the 3rd and the 4th ring road; otherwise 0
ringRoad45	1 if the complex is between the 4th and the 5th ring road; otherwise 0
ringRoad56	1 if the complex is between the 5th and the 6th ring road; otherwise 0
ringRoad600	1 if the complex is out of the 6th ring road; otherwise 0
4.Month Dummies	
Month $2 \sim 36$	Month dummy; 1 if the transaction is in this month, otherwise 0

Table 3.3. Summary Statistics.

Variable	Mean	Std. Dev.	Min.	Max.
1.Unit-level attributes				
lnprice	5.842	0.729	1.308	9.952
Bedrooms	2.262	1.019	0	21
DiningRooms	1.441	0.63	0	7
area	111.216	77.282	6	2600
floor	14.809	9.146	0	102
Villa	0.067	0.251	0	1
West	0.036	0.186	0	1
East	0.047	0.212	0	1
South	0.162	0.369	0	1
Northwest	0.027	0.162	0	1
Northeast	0.03	0.17	0	1
Eastwest	0.028	0.165	0	1
Southwest	0.057	0.233	0	1
Southeast	0.074	0.262	0	1
Southnorth	0.495	0.5	0	1
2.Complex-level characteristics				
Green	0.349	0.077	0.002	0.9
FAR	2.515	1.236	0.04	10.7
WaterView	0.015	0.121	0	1
HighRise	0.736	0.441	0	1
3.Geographic attributes				
NearSubway	0.365	0.481	0	1
NearPark	0.045	0.208	0	1
ringRoad02	0.049	0.216	0	1
ringRoad23	0.121	0.326	0	1
ringRoad34	0.228	0.42	0	1
ringRoad45	0.255	0.436	0	1
ringRoad56	0.27	0.444	0	1
4.Monthly Sold Dummies[2]				
Month 2	0.008	0.087	0	1
Month 3	0.021	0.145	0	1
Month 4	0.014	0.119	0	1
Month 5	0.023	0.151	0	1
N		190580		

[2] Only 4 monthly sold dummies are listed here. For the left 31 monthly sold dummies, see table A.1.

Table 3.4. Results of OLS and Quantile Regression.

	OLS	Q(0.10)	Q(0.20)	Q(0.30)	Q(0.40)	Q(0.50)	Q(0.60)	Q(0.70)	Q(0.80)	Q(0.90)
Bedrooms	0.0975***	0.107***	0.0880***	0.0759***	0.0633***	0.0520***	0.0398***	0.0305***	0.0208***	0.0156***
DiningRooms	0.145***	0.132***	0.119***	0.116***	0.119***	0.120***	0.116***	0.112***	0.100***	0.0866***
area	0.0049***	0.0049***	0.0059***	0.0064***	0.0067***	0.0070***	0.0073***	0.0075***	0.0077***	0.0079***
floor	0.0030***	0.0066***	0.0054***	0.0040***	0.0031***	0.0022***	0.00159***	0.00083***	0.0004***	0.0002
Villa	0.0116***	-0.0291***	-0.0933***	-0.110***	-0.102***	-0.0876***	-0.0763***	-0.0747***	-0.0962***	-0.0933***
West	0.0399***	0.0428***	0.0477***	0.0546***	0.0381***	0.0519***	0.0507***	0.0467***	0.0404***	0.0484***
East	0.0581***	0.0732***	0.0768***	0.0733***	0.0621***	0.0629***	0.0677***	0.0638***	0.0571***	0.0485***
South	0.0860***	0.109***	0.103***	0.103***	0.0912***	0.0960***	0.105***	0.105***	0.0943***	0.0805***
Northwest	0.0753***	0.132***	0.105***	0.0977***	0.0845***	0.0790***	0.0692***	0.0591***	0.0488***	0.0169*
Northeast	0.0735***	0.128***	0.105***	0.106***	0.0887***	0.0873***	0.0779***	0.0648***	0.0417***	0.0046
Eastwest	0.176***	0.192***	0.185***	0.184***	0.177***	0.183***	0.181***	0.157***	0.139***	0.124***
Southwest	0.136***	0.166***	0.141***	0.134***	0.118***	0.123***	0.120***	0.110***	0.0990***	0.0823***
Southeast	0.141***	0.165***	0.138***	0.135***	0.124***	0.129***	0.126***	0.116***	0.103***	0.0861***
Southnorth	0.210***	0.208***	0.189***	0.189***	0.184***	0.194***	0.198***	0.189***	0.175***	0.151***
Green	0.471***	0.328***	0.327***	0.331***	0.360***	0.353***	0.332***	0.349***	0.378***	0.438***
FAR	0.0016**	0.0075***	0.0073***	0.0063***	0.0038***	-0.0003	-0.0046***	-0.0098***	-0.0162***	-0.0189***
WaterView	0.158***	0.108***	0.0938***	0.0662***	0.0376***	0.0297**	0.0310**	0.0153*	-0.0006	-0.0192**
HighRise	0.0766***	0.0807***	0.0889***	0.0976***	0.101***	0.103***	0.0929***	0.0786***	0.0539***	0.0007
NearSubway	0.0376***	0.0494***	0.0487***	0.0412***	0.0311***	0.0237***	0.0228***	0.0229***	0.0217***	0.0207***
NearPark	0.0222***	0.0929***	0.0618***	0.0378***	0.0242***	0.0121***	0.0087**	0.0103***	0.0102***	0.0057
ringRoad02	1.501***	1.745***	1.679***	1.648***	1.621***	1.565***	1.502***	1.400***	1.275***	1.251***
ringRoad23	1.340***	1.573***	1.513***	1.496***	1.477***	1.424***	1.364***	1.259***	1.126***	1.094***
ringRoad34	1.329***	1.564***	1.496***	1.470***	1.442***	1.385***	1.329***	1.228***	1.100***	1.072***
ringRoad45	1.188***	1.427***	1.371***	1.352***	1.328***	1.263***	1.194***	1.078***	0.939***	0.903***
ringRoad56	0.866***	1.071***	1.020***	0.992***	0.972***	0.930***	0.898***	0.820***	0.709***	0.702***
N	190580	190580	190580	190580	190580	190580	190580	190580	190580	190580

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The basic OLS regression results are shown in the second column of table 3.4. The coefficients of dummy variables is a measure of difference between the reference group and the base group, and the difference can be calculated accurately as $e^{coefficient} - 1$. OLS estimates show that all coefficients are statistically significant at 1% level. All the coefficients have the expected sign except FAR. FAR is the floor area ratio, which is defined as the total square meter of a development divided by total square meter of the land area the building

Table 3.5. Coefficients of Monthly Sold Dummies.

	OLS	Q(0.10)	Q(0.20)	Q(0.30)	Q(0.40)	Q(0.50)	Q(0.60)	Q(0.70)	Q(0.80)	Q(0.90)
Month2	0.0490***	0.0462**	0.0729***	0.0462***	0.0411***	0.0518***	0.0560***	0.0564***	0.0541***	0.0468***
Month3	0.0815***	0.0892***	0.0923***	0.0824***	0.0787***	0.0930***	0.0962***	0.0892***	0.0905***	0.0779***
Month4	0.109***	0.141***	0.140***	0.117***	0.102***	0.120***	0.120***	0.115***	0.111***	0.0939***
Month5	0.140***	0.149***	0.162***	0.152***	0.152***	0.157***	0.161***	0.144***	0.126***	0.0974***
Month6	0.164***	0.172***	0.182***	0.167***	0.162***	0.178***	0.179***	0.161***	0.146***	0.131***
Month7	0.165***	0.183***	0.189***	0.172***	0.162***	0.169***	0.174***	0.162***	0.151***	0.144***
Month8	0.176***	0.203***	0.212***	0.190***	0.176***	0.188***	0.182***	0.168***	0.157***	0.138***
Month9	0.203***	0.230***	0.244***	0.225***	0.211***	0.221***	0.216***	0.192***	0.168***	0.147***
Month10	0.197***	0.242***	0.249***	0.227***	0.208***	0.216***	0.210***	0.187***	0.169***	0.152***
Month11	0.214***	0.259***	0.261***	0.241***	0.225***	0.228***	0.224***	0.200***	0.188***	0.163***
Month12	0.221***	0.262***	0.265***	0.244***	0.226***	0.230***	0.219***	0.196***	0.186***	0.169***
Month13	0.224***	0.266***	0.269***	0.250***	0.230***	0.230***	0.224***	0.207***	0.191***	0.173***
Month14	0.227***	0.260***	0.271***	0.256***	0.238***	0.242***	0.232***	0.212***	0.197***	0.176***
Month15	0.222***	0.248***	0.263***	0.247***	0.224***	0.232***	0.225***	0.208***	0.195***	0.181***
Month16	0.214***	0.233***	0.251***	0.234***	0.222***	0.224***	0.219***	0.204***	0.193***	0.178***
Month17	0.184***	0.213***	0.219***	0.202***	0.191***	0.191***	0.187***	0.171***	0.154***	0.139***
Month18	0.161***	0.165***	0.187***	0.179***	0.171***	0.174***	0.169***	0.148***	0.135***	0.117***
Month19	0.138***	0.154***	0.171***	0.156***	0.143***	0.149***	0.146***	0.129***	0.116***	0.0962***
Month20	0.109***	0.135***	0.146***	0.128***	0.115***	0.122***	0.118***	0.101***	0.0939***	0.0844***
Month21	0.138***	0.151***	0.165***	0.157***	0.148***	0.157***	0.147***	0.125***	0.111***	0.105***
Month22	0.146***	0.152***	0.172***	0.161***	0.157***	0.162***	0.152***	0.131***	0.124***	0.109***
Month23	0.161***	0.156***	0.176***	0.168***	0.165***	0.175***	0.170***	0.145***	0.138***	0.123***
Month24	0.155***	0.145***	0.181***	0.179***	0.167***	0.175***	0.161***	0.143***	0.141***	0.126***
Month25	0.141***	0.115***	0.168***	0.177***	0.167***	0.166***	0.164***	0.146***	0.134***	0.114***
Month26	0.155***	0.136***	0.180***	0.185***	0.177***	0.177***	0.166***	0.157***	0.150***	0.128***
Month27	0.153***	0.128***	0.180***	0.179***	0.169***	0.174***	0.165***	0.155***	0.143***	0.126***
Month28	0.183***	0.160***	0.203***	0.204***	0.195***	0.204***	0.202***	0.177***	0.164***	0.150***
Month29	0.184***	0.145***	0.197***	0.200***	0.200***	0.212***	0.204***	0.184***	0.174***	0.176***
Month30	0.198***	0.162***	0.210***	0.202***	0.206***	0.214***	0.206***	0.188***	0.188***	0.185***
Month31	0.206***	0.175***	0.216***	0.224***	0.225***	0.230***	0.228***	0.204***	0.196***	0.188***
Month32	0.200***	0.168***	0.207***	0.217***	0.214***	0.222***	0.225***	0.208***	0.203***	0.198***
Month33	0.195***	0.148***	0.186***	0.195***	0.205***	0.217***	0.220***	0.211***	0.211***	0.205***
Month34	0.185***	0.171***	0.195***	0.191***	0.192***	0.208***	0.205***	0.197***	0.202***	0.203***
Month35	0.199***	0.172***	0.219***	0.226***	0.220***	0.225***	0.216***	0.190***	0.199***	0.207***
Month36	0.198***	0.149***	0.185***	0.201***	0.207***	0.225***	0.226***	0.217***	0.236***	0.245***
Constant	3.175***	2.509***	2.672***	2.801***	2.930***	3.075***	3.248***	3.488***	3.771***	3.991***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

is located on. FAR is not common in typical Western studies while it is relevant to high-density building market like China(see Liao and Wang, 2012). Higher FAR indicate more dense housing construction. So the sign of FAR is expected to be negative. However, the coefficient of FAR is small compared with other attributes and there is also quantile effect that we will discuss later.

The estimated coefficients of variables Bedrooms, DiningRoooms, area, and floor can be interpreted as the percentage increase of the corresponding housing attributes to conditional transaction price. House price increases as area and floor of a housing unit increase. Holding everything else constant, one more bedroom and dinning room increase house price9.75% and 14.5%, respectively. Price is 1.17 percent higher for villa compared with non-villa unit, ceteris paribus. Houses facing west, east, and south are more valuable compared with the comparison group that houses facing north. House price is even higher for the units facing two directions, for example, there are 23.37% increase for houses facing both south are north compared with houses only face north. Also, houses facing both east and west(19.24%) are more valuable than houses that face either east(5.98%) or west(4.07%). Since most of the houses in China are condominium units, a house faces two directions will have more chance to get sunshine and thus has a more comfortable living condition. The premium of housing units that face both north and south may explain why almost half of the housing units (49.5%) in the data set face both south and north and also why this kind of house is most popular in China.

The estimates also show that house prices are expected to be 47.1 percent higher if the greening rate increase by 100%. People do prefer properties with water view and easy accessibility to a park: the availability of water view and park increase the mean value of house prices by 17.12% and 2.24%, respectively. Apartments in high-rise development are 7.96% higher by mean value, ceteris paribus. Home-buyers also favor complexes that are close to a subway station. The estimated mean premium for the convenience of the subway is 3.83%. Since the traffic congestion problem becomes more and more severe in big cities like Beijing, the demand for subway proximity of housing is rising in China. Xu et al. (2015) estimates the demand of subway after private driving restrictions were imposed by the Beijing City government in 2008. They find that people are willing to pay 1.8% to 2.7% more for houses with subway convenience.

We can see that the ring road dummy variables all contribute to the house prices significantly while the coefficients decreases indicating that the house price decreases as the development is farther away from the city center. The house price is 1.5 times higher when a development is within the second ring road compared with the reference group, ceteris paribus. The coefficient of ringRoad56 reveals that the houses between the fifth ring road and the sixth ring road are 86.6% higher than the reference group. The results are reasonable since the resources are distributed unequally in Beijing. As the house is closer the city center, some resources such as hospitals, quality schools and work opportunities are more accessible for households.

In addition, we examine home buyers' marginal willingness to pay for a housing attribute across the full distribution of house price, and find that lower-priced home buyers value most housing attributes differently from higher-priced home buyers. We estimate 9 representative quantiles from 0.1 through 0.9 with an increment of 0.1, which we denote as $Q(0.10)$, $Q(0.20)$, $\cdots$ $Q(0.90)$. The results of quantile regression are reported in the last 9 columns of Table 3.4. To give some further insights into the evolvement of coefficients, scatter plots of the coefficients by quantile are shown in Figure 3.1.

As can be observed from the regression results and scatter plots, several variables exhibit significant quantile effects. The estimated coefficients of quantile regression for some variables differ from the OLS estimates and generally vary across distributions. The coefficient of bedrooms shows a clear downward pattern. The premium associated with an additional bedroom decrease from – 0.107 at quantile 0.10 to – 0.016 at quantile 0.90. The result is partly consistent with the finding of Liao and Wang (2012), in which the coefficient of bedrooms also shows a downward pattern from lower quantiles to middle quantiles. The

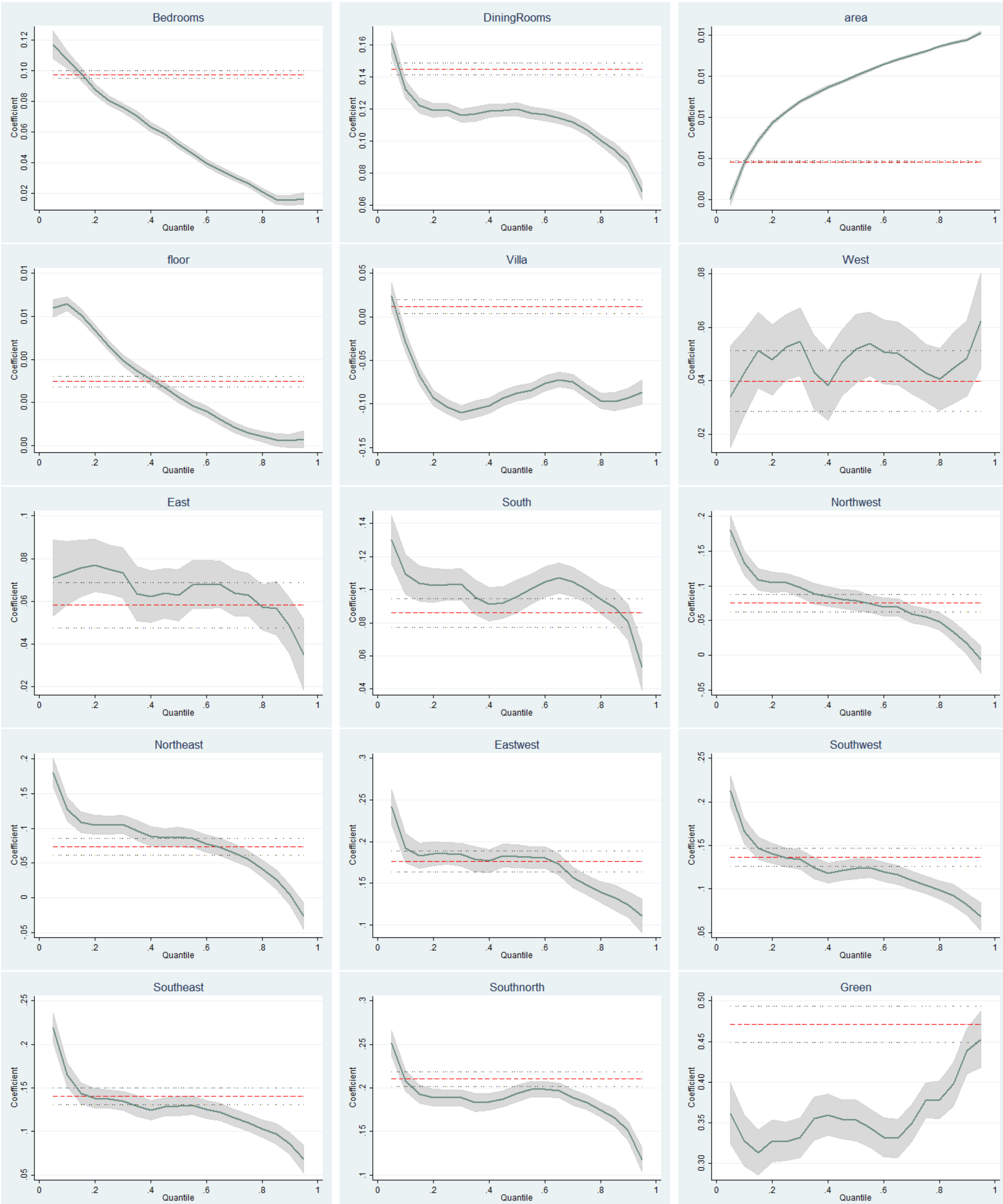

Figure 3.1. Coefficient Estimates by Quantile.

result is reasonable since lower-priced house consumers may be more inclined with a big family in China, thus, have a higher demand for more bedrooms. For the living rooms, the

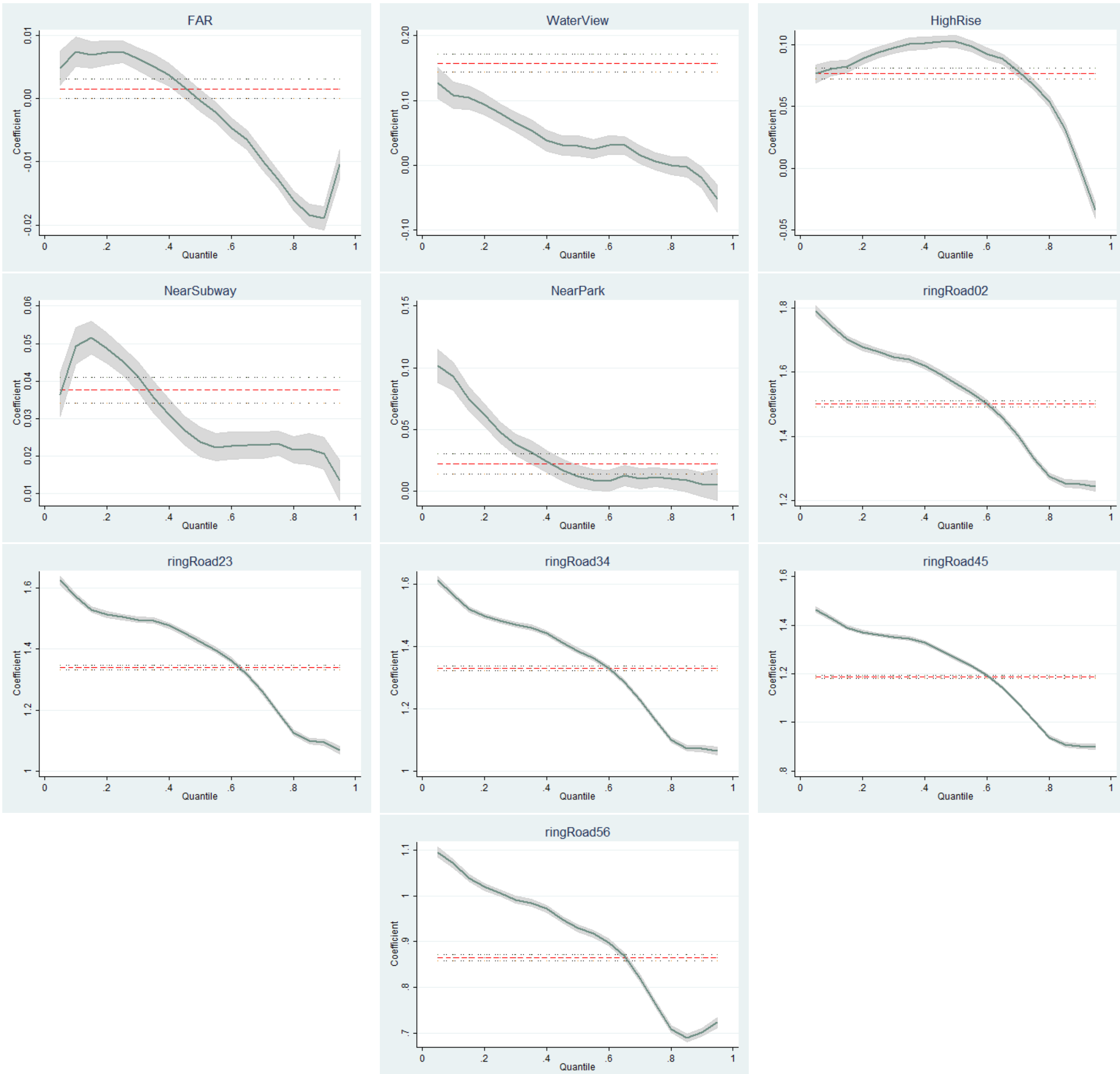

Figure 3.1. Coefficient Estimates by Quantile (Continued).

pattern is similar: there is a sharp decrease of the coefficient at the lower quantiles, then the coefficient remains mostly stable from quantile 0.2 to quantile 0.7, finally following a slow decrease as the quantile increases to 0.8 and 0.9. An additional unit of floor area adds more value to house price as the quantile increases (For example, 0.0079% at quantile 0.9 VS 0.0049% for at quantile 0.1). Increase by floor number depicts a reverse tendency as house price increase from lower quantiles to upper quantiles. All of the coefficients for floor number are significant at 1% level except quantile 0.9, which is not significant even at 10% level.

The coefficients of villa seem strange as the OLS coefficient is positive while most quantile coefficients are negative, positive coefficient only appears at the lowest quantile. The quantile effects of the coefficients of direction attributives demonstrate downward pattern for all except west. There is no explicit trend across quantiles for the house facing west.

Quantile effects are evident for greening rate and FAR, even though different patterns come into sight. The beneficial effects of greening rate are much more obvious at higher quantiles that seems high-price house owners are more concerned with the living environment. We find, from lower quantiles to higher quantiles, the coefficients of FAR decrease from positive to negative and a sharp increase but still negative at extreme higher quantiles. The coefficients of WaterView shows a flat decreasing trend as quantile of house price increases. The quantile coefficient of Green and WaterView is all lower below the OLS estimates. We can see that the estimates of HighRise are positive and rise slowly from lower quantile to middle quantile, then following an increasing rate of decrease resulting into negative coefficients at higher quantiles.

Near a subway station appeared to add more values to house prices at lower quantiles because the coefficient decreases from lower quantiles to higher quantiles. This may reflect that high-price house consumers are more possible to own cars that decrease their demand for subway transportation. In Beijing, not every family has a car especially for low income households, so the demand for subway proximity is stronger for low-price house buyers. We can see low-priced house buyers are more concerned with whether the complex is near a park than high-priced house buyers. The reason may be that high-priced house is more likely to locate in places with better environment. Thus, they may not have a strong demand for park sightseeing as low-price house owners.

For a series of dummy variables ringRoad, almost the same pattern appears. The coefficients decrease from lower quantiles to upper quantiles. For instance, the house within the second ring road increase house prices by more percentage at lower price quantiles (174.5%

at quantile 0.1 while 125.1% for at quantile 0.9) compared with houses outside of the sixth ring road in Beijing. All the coefficients for variables Ringroad are significant at 1% level.

3.5.2. House Price Indices

Our primary goal is to see whether there are some variations on the appreciation rate of house price across the full distribution of house price in Beijing. So we estimate house price indices from time dummies by employing OLS and quantile regression model. House price indices are monthly based, from January 2013 to December 2015. The index of the base period January 2013 is 1. Figure 3.2 shows the evolvement of house price indices.

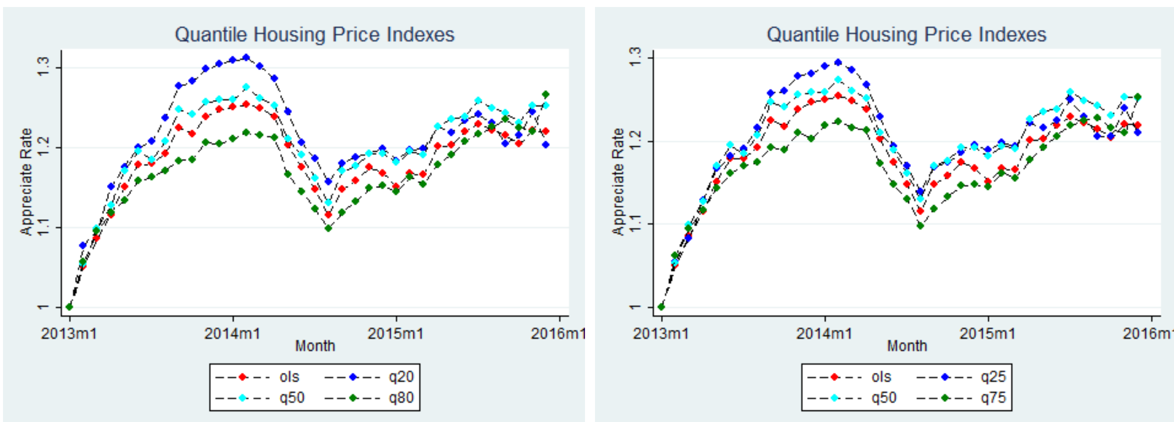

Figure 3.2. Quantile Housing Price Indexes.

House price indices estimated by both ordinary least square and quantile regression models increased from January 2013 to February 2014, where the index peaked at all quantiles during the study period. In February 2014, the index of summit calculated by OLS, quantile regression at quantile 0.20, quantile 0.50, and quantile 0.80 is 1.254, 1.311, 1.274, and 1.218, respectively. A large difference of 0.093 between quantile 0.20 and quantile 0.80 appears in February 2014. The magnitude of appreciation rates of house price suggested by OLS is greater than the rates of higher quantile but lower than the resulting from lower quantile. The median percentage index is close to the OLS index during the period of January 2013 to February 2014.

Following the hot housing market year of 2013, there is a sudden decline for the whole distribution of house price from February 2014 to August 2014. For example, the house price index at quantile 0.2 dropped from 1.331 to 1.157 while the OLS index fell down from 1.254 to 1.115. On quantile 0.50 and quantile 0.80, the index decreased from 1.274 to 1.130 and from 1.218 to 1.098, respectively. The difference of house price index between quantile 0.2 and quantile 0.8 is 5.9 percent, which means after a period of housing market recession, the gap of index between lower quantile and higher quantile narrows. We can also see a delay of house price decline on higher quantiles compared with middle and lower quantiles. For instance, at quantile 0.80 and quantile 0.75, the decreases of index from February 2014 to April 2014 is only 0.005 and 0.01, respectively. While at quantile 0.5 and quantile 0.20, the decreases of the index during the same period are 0.015 and 0.025, respectively. After April 2014, the house price at higher quantiles started to drop faster than at lower quantiles.

After August 2014, the house prices enter into a comparing slower appreciation process until December 2015, where the house price is still slightly lower than the peak reached in February 2014. The monthly growth rate of indices at high quantiles speed up and surpass low quantiles after August 2014, resulting that the gap between low percentages and high percentages decrease even disappears in December 2015. The disparity of house price indices among different quantiles even narrowed after August 2014. For example, from August 2014 to January 2015, the house price index at quantile 0.8 increased from 1.098 to 1.144 coincidence with the index at quantile 0.2 rose from 1.157 to 1.183. Then the distance of index between quantile 0.2 and quantile 0.8 became 0.039.

Moreover, even though house price indices across quantiles seem to share similar temporal patterns, there is substantial variation on the appreciation rate of house prices among different quantiles of house price distributions. The total appreciation rate is significantly lower for higher quantiles than lower quantiles of distributions across almost all periods from January 2013 to December 2015.

3.6. Conclusion and Discussion

The China's decade-long housing boom from 2000 has gained considerable interests and concerns especially after the growth speed of China's economy seems to slow down since 2008. Due to China's rising importance in the world economy, building a high-quality and in-depth house price index is of crucial importance to detect the potential housing bubble, measure housing affordability, and evaluate economic risk.

Often, micro-level data are not available in nascent markets such as China and reliability of data are also under suspicion. However, as information technology develops in China, housing data become accessible in Chinese housing websites. Utilizing recent data mining technology, we collected sequential transaction data of new homes within the same housing developments in Beijing from January 2013 through December 2015 in a leading Chinese housing website.

From the perspective of methodology, our method of data mining provides an attempt of a new approach to collect data. Researchers can assemble the preference and decision-making details of consumers using the new technologies, which may provide new insight to better understand the sweep of economic trends, especially for developing countries that may lack quality data.

Moreover, our primary goal is to study the housing market in Beijing to get a quality house price index. Utilizing by far the most comprehensive public data for newly-built residential housing units, this study provides insight into the evolution of China's house price distributions. Using transaction data and detailed characteristics of developments, we employ a quantile regression approach on hedonic price model to construct house price indices across the conditional distribution of the house price. We control housing unit-level characteristics such as number of bedrooms, number of living rooms, square meters of living area, orientation of the house, etc.; housing complex-level characteristics such as green space rate, floor area ratio, etc.; and geographic attributes such as close to subway, park, etc.

Two main findings stand out in our study. First, there is substantial appreciation of house price–17.6% to 27.1% for different quantiles–from January 2013 to February 2014, where the house prices peak for all quantiles during our research period. Following that there is a sudden decline for the whole distribution of house price from February 2014 to August 2014. After that the house prices enter into a comparing slower appreciation process until December 2015, where the house price is still slightly lower than the peak reached in February 2014. Second, even though house price indices across quantiles seem to share similar temporal patterns, there is substantial variation on the appreciation rate of house prices among different quantiles of house price distribution. The appreciation rate is significantly lower for higher quantiles of price distribution than the lower quantiles of distribution across almost all periods from January 2013 to December 2015. In addition, we examine home buyers' marginal willingness to pay for a housing attribute, and find that lower-priced home buyers value most housing attributes differently from higher-priced home buyers.

4. DECOMPOSITION

4.1. Introduction

China has witnessed rapid urbanization and concurrently real estate boom in the last decade. The over decade-long housing boom from 2000 have gained considerable interests and concerns especially after the growth speed of China's economy seems to slow down since the financial crisis of 2008. But, house price in China also increased substantially after the 2008 financial crisis, especially in large cities. Estimated by Fang et al. (2015), in first-tier cities, which include the four most populated and most economically important metropolitan areas in China–Beijing, Shanghai, Guangzhou, and Shenzhen, house prices had an average annual real growth rate of 13.1 percent from 2003 to 2013. Using recent transaction data of newly-built housing units, it is also estimated that the nominal house price on average rose by roughly 51% between 2012 and 2015 in Beijing.

However, house price appreciation may not be shared equally by different houses. Some houses may appreciate rapidly while other homes may remain stable or even depreciated. For example, low-level houses may depreciation if there is substantial suburbanization but housing units in the upper segment may appreciate rapidly in the case of upgrading housing characteristics, leading to heterogeneity in price gaps across the house price distribution. We find that there is depreciation in the lowest segment of distribution between 2012 and 2015 but housing units appreciate rapidly in the higher segment, resulting in a thicker distribution on both the left and right in 2015. With this background in mind, we want to examine house price changes in the full distribution. Specifically, one question we want to ask is why do house prices changed unequally in different locations of the distribution in Beijing. The answer to this question is still not clear because price changes may be ascribed to variations in the characteristics of housing units. For instance, entry-level and high-priced housing units were sold more frequently in 2015 than in 2012. Or alternatively, distributional changes may be not driven by the housing characteristics but by changing returns to housing characteristics in the underlying hedonic pricing functions. In this chapter, we investigate this question using a mean decomposition approach developed by Oaxaca (1973) and Blinder (1973), and a quantile decomposition method developed by Chernozhukov et al. (2013), respectively.

There is no publicly accessible real estate transaction dataset that contains comprehensive housing attributes released by any statistical authority yet, but the transaction data can be found on a set of real estate websites in China. So we obtained and assembled a detailed housing dataset using a web scraper programmed by python from one of the largest housing website "Soufun"[1]. The dataset contains transactions of house sales for year 2012 and 2015. We observe that house price appreciated by roughly 51% between 2012 and 2015.

OLS regression and quantile regression are estimated first. The regression results show how a particular housing attribute is priced implicitly on average or at specific positions on the conditional distribution of the house price. Either the OLS regression or quantile regression results suggest that housing characteristics have great power in explaining house prices in both years. Conditional quantiles obtained from quantile regression provide the basis to implement the decomposition method of Chernozhukov et al. (2013). Once the conditional quantiles become available, counterfactual distribution of house price in 2015 can be calculated by re-weighting the characteristics of houses in 2012 (period 0) so that they look like the characteristics in 2015 (period 1), holding that the conditional distribution of house price in 2015 is fixed. Then, we can carry out the decomposition analysis to distinguish changes in the distribution of house price over time into two components: one component is the so-called composition effect, which captures changes driven by variations of housing characteristics between 2012 and 2015; the other component is the coefficient effect, which is attributed to varying implicit prices in the underlying hedonic functions.

[1] The address of Soufun is http://www.fang.com/

The results of Oaxaca-Blinder mean decomposition show that altered implicit returns to housing characteristics rather than the variations of housing characteristics play a dominant role in explaining the gap of the mean house price between 2012 and 2015. Specifically, only 13% of the price gap is attributed to the temporal differential of housing characteristics while the other 87% are caused by varying implicit prices. However, quantile decomposition provides additional evidence by taking the full distribution of house prices into consideration. Quantile decomposition results suggest that gaps in house prices that can be attributed to the composition effect rise monotonously from negative in the left tail of the distribution to positive in the right end, the while gaps that can be explained by the coefficient remain stable, only slightly different from the mean. Thus, the proportion of price gaps that are explained by the composition effect increase monotonously, leading to even greater role play by the composition effect rather than the coefficient effect. Our findings suggest that quantile analysis supplement information of heterogeneity in the role of housing characteristics in the determination of house price distribution that means analysis cannot provide, either regression or decomposition.

Decomposition analysis has been widely used in labor economics, especially in studies of wage inequality (e.g. Buchinsky, 1994; Gosling et al., 2000; Machado and Mata, 2005a; Melly, 2005) since it was introduced by Oaxaca (1973) and Blinder (1973), Recently, a growing number of literature have focused on housing related areas, and use quantile decomposition methods to explain the differential in the distribution of house prices or rental prices or land prices between two periods or two groups (McMillen, 2008; Cobb-Clark and Sinning, 2011; Nicodemo and Raya, 2012; Fesselmeyer et al., 2013; Thomschke, 2015; Qin et al., 2016). McMillen (2008) decomposes temporal changes in the distribution of house price from 1995 to 2005 in Chicago using the method developed by Machado and Mata (2005a). McMillen (2008) finds high-price houses in Chicago experienced a higher rate of appreciation, leading to a less skewed distribution of house price in 2005. Moreover, decomposition results of McMillen (2008) suggest that varying quantile regression coefficients over time rather than the temporal changes in house characteristics are responsible for distribution changes in house price. The method of Machado and Mata (2005a) has also been applied by Fesselmeyer et al. (2013) to analyze the white-black house value gap from 1997 to 2005. Unlike the finding of McMillen (2008), Fesselmeyer et al. (2013) present evidence that altered housing characteristics of white- and black-owned houses play a dominant role in explaining the racial gap. Nicodemo and Raya (2012) focus on the housing market in Spain, using the approach of Melly (2005) to examine changes in the distribution of house price in six cities of Spain from 2004 to 2007. Thomschke (2015) use the method developed by Chernozhukov et al. (2013) to study changes in the distribution of rental prices in Berlin from 2007 to 2012. The decomposition results of Nicodemo and Raya (2012) and Thomschke (2015) are in line with McMillen (2008), that is, the temporal changes in the distribution of house price or rental price are mainly driven by the differences in returns to housing characteristics rather than the variation of characteristics over time. While most of housing related studies have focused on the housing market in developed countries, litter attention has been paid to nascent housing markets in developing countries like China. Qin et al. (2016) provide an attempt to analyze the temporal gaps in the price distribution of different types of land in urban China from 2007 to 2012. Decomposition results of Qin et al. (2016) indicate that for different types of land, the proportion of price gaps that can be explained by the composition effect or the coefficient varying with the position of the price distribution.

Although the findings of these studies appear to be not consistent in terms of the contributions of composition effect and coefficient effect, the heterogeneity in the decomposition results is partly due to the diversity of housing markets. For example, a nascent housing market like China is experiencing substantial structural changes during a short period while the housing market in developed countries may remain stable in a relatively long period since not many housing units are constructed or upgraded. In this sense, our study is of value as we contribute to the literature by providing evidence from the newly-built housing market, though the method used in this research is closely related to previous studies.

Unsurprisingly, we observe substantial variations on the housing characteristics over time in Beijing, and further decomposition results suggest that even the contribution to house price gap attributed to changes in the housing characteristics are lower than the contribution caused by varying coefficient at most of positions in the distribution or on average, the relative contribution of the composition effect are comparable to the coefficient effect. Also, the contribution of the composition effect is greater than similar studies of what McMillen (2008) and Nicodemo and Raya (2012) find. In addition, our research provides policy implications in terms of housing conditions in Beijing. That is, regarding housing characteristics, the living condition of low-price house buyers were worse in 2015 even house prices appreciated, while the living condition of high-price consumers was better in 2015 but at the expense of relatively higher payment than in 2012.

The rest of this chapter proceeds as follows. Section 4.2 introduces the empirical methodologies used in this chapter. We discuss the data set in Section 4.3. Section 4.4 summarizes main empirical results of regression and decomposition of house price distributions. Section 4.5 provides some discussion and concludes.

4.2. Empirical Methodology

We are interested in the determinations of changes on house prices across two periods. Specifically, we want to know to what extent the increase in house price from 2012 to 2015 in Beijing was caused by the variations of the characteristics, and the portion which can be attributed to changes in the coefficients. In this regarding, decomposition methods that are widely used to study group differentials (such as race, sex, and so on) in the labor market are appropriate. One well-known approach is Oaxaca-Blinder decomposition, which is attributed by the work of Oaxaca (1973) and (Blinder, 1973). Both work divide the wage differential between two groups into one portion that can be quantified by group differences in characteristics such as education or experience, and a part that cannot be explained by such differences in the determination of labor income.

4.2.1. Oaxaca-blinder Decomposition

Following conventional practice in housing-related literature, we assume that natural logarithm of transaction price of a house, Y, is a function of a vector of housing attributes, $\mathbf{X}$. There are two periods 2012 and 2015, thus

$$Y_t = \mathbf{X}'\boldsymbol{\beta}_t + \epsilon_t,\ for\ t\ =\ 12,\ 15 \tag{4.1}$$

where ϵ_t the error term and $E(\epsilon_t|\mathbf{X}) = 0$. Let I_{2015} be an indicator of the period 2015. By taking expectations over $\mathbf{X}$, the overall mean house price difference $\triangle Y$ can be written as

$$\begin{aligned}
\triangle Y &= E(Y_{15}|I_{15} = 1) - E(Y_{12}|I_{15} = 0) \\
&= E[E(Y_{15}|\mathbf{X}', I_{15} = 1)|I_{15} = 1] - E[E(Y_{12}|\mathbf{X}', I_{15} = 0)|I_{15} = 0] \\
&= [E(\mathbf{X}'|I_{15} = 1)\boldsymbol{\beta}_{15} + E(\epsilon_{15}|I_{15} = 1)] - [E(\mathbf{X}'|I_{15} = 0)\boldsymbol{\beta}_{12} + E(\epsilon_{12}|I_{15} = 0)]
\end{aligned}$$

where $E(\epsilon_{15}|I_{15} = 1)]$ and $E(\epsilon_{12}|I_{15} = 0)]$ both is zero. Adding and subtracting the mean counterfactual price that houses of period 2012 would be valued under the coefficients of period 2015, $E(\mathbf{X}|I_{15} = 0)\boldsymbol{\beta}_{15}$, the equation above becomes

$$\begin{aligned}
\triangle Y &= E(\mathbf{X}'|I_{15} = 1)\boldsymbol{\beta}_{15} - E(\mathbf{X}'|I_{15} = 0)\boldsymbol{\beta}_{15} \\
&\quad + E(\mathbf{X}'|I_{15} = 0)\boldsymbol{\beta}_{15} - E(\mathbf{X}'|I_{15} = 0)\boldsymbol{\beta}_{12} \\
&= \boldsymbol{\beta}_{15}(E(\mathbf{X}'|I_{15} = 1) - E(\mathbf{X}'|I_{15} = 0)) + E(\mathbf{X}'|I_{15} = 0)(\boldsymbol{\beta}_{15} - \boldsymbol{\beta}_{12}) \\
&= \triangle Y^x + \triangle Y^u
\end{aligned}$$

The decomposition can be estimated by replacing the expected values $E(\mathbf{X}'|I_{15} = 0)$ and $E(\mathbf{X}'|I_{15} = 1)$ by sample average $\bar{\mathbf{X}}_{12}$ and $\bar{\mathbf{X}}_{15}$, the expression then becomes

$$\begin{aligned}
\hat{\triangle} Y &= \bar{\mathbf{X}}'_{15}\hat{\boldsymbol{\beta}}_{15} - \bar{\mathbf{X}}'_{12}\hat{\boldsymbol{\beta}}_{15} + \bar{\mathbf{X}}'_{12}\hat{\boldsymbol{\beta}}_{15} - \bar{\mathbf{X}}'_{12}\hat{\boldsymbol{\beta}}_{12} \\
&= \hat{\boldsymbol{\beta}}_{15}(\bar{\mathbf{X}}'_{15} - \bar{\mathbf{X}}'_{12}) + \bar{\mathbf{X}}'_{15}(\hat{\boldsymbol{\beta}}_{15} - \hat{\boldsymbol{\beta}}_{12}) \\
&= \hat{\triangle} Y^x + \hat{\triangle} Y^u
\end{aligned}$$

where the first term $\hat{\triangle}Y^x$ captures the mean movement in house price that is determined by difference on the housing characteristics across the two periods, which is usually called the "composition" or "characteristic" effect in housing related literature, the second term $\hat{\triangle}Y^u$ is the "coefficient" effect, which measure the change of house price that is attributed to change in hedonic coefficients. While in some other areas like wage determination in labor economics, the second part has also been called the "unexplained" part of the wage differentials and is commonly attributed to "discrimination" of gender or race.

4.2.2. Quantile Decomposition

Even though the Oaxaca-blinder decomposition has been applied extensively in empirical economics studies, its insight is limited to the mean of interested outcome. However, the value in a decomposition analysis can be any location at the distribution of Y, such as 10 percentile, median, or 90 percentile, i.e., the decomposition estimation is not restricted on mean values while can be extended to different positions of the entire distribution. A great number of methods aimed at extending and refining the Oaxaca-blinder decomposition beyond the mean, and to further explain the difference in full distribution have been proposed.(e.g., DiNardo et al. (1996); Gosling et al. (2000); Machado and Mata (2005b); Melly (2006); Firpo et al. (2009); Chernozhukov et al. (2013), among others.)[2] Chernozhukov et al. (2013) provide estimation and moreover inference procedures for the entire marginal distribution of house price and its functions based on a series of regression methods such as distribution regression and quantile regression.

Quantile regression (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001) allows housing attributes to be valued differently at multiple locations of the house price distribution: the estimated coefficients reflect the consumers' willingness to pay for each attribute at different quantiles. Quantile regression defines an objective function that is weighted asymmetrically, of which the weights 2ρ and $2-2\rho$ are assigned to the observations depending on whether the residual is above or below zero, for the selected ρth quantile. The coefficients

[2]see Fortin et al. (2011) for a comprehensive review of decomposition methods.

are obtained by minimizing the following objective function

$$\min \sum_i \{2\rho - 2 \times 1(Y_i \leq \mathbf{X}_i'\boldsymbol{\beta})\}|Y_i \leq \mathbf{X}_i'\boldsymbol{\beta}| \tag{4.2}$$

where the weight takes value 2ρ when the residual $y_i \leq \mathbf{X}_i'\boldsymbol{\beta}$ is negative and $2\rho - 2$ otherwise. For instance, to estimate the 25th quantile regression, the positive residuals are weighted by $2\rho = 0.5$ while the negative residuals are weighted by $2 - 2\rho = 1.5$. Only at the median the objective function is weighted symmetrically, i.e., $2\rho = 2 - 2\rho = 1$, and the objective function can be simplified as

$$\min \sum_i |Y_i \leq \mathbf{X}_i'\boldsymbol{\beta}| \tag{4.3}$$

Just like the median splits the sample into two equal subgroup, the median regression divides the residuals equally into half above the estimated equation while the other half below.

We follow the procedure developed by Chernozhukov et al. (2013) to decompose the changes of house prices from 2012 to 2015 over the entire distribution. In order to do that, the quantile regression coefficients are estimated in the first stage.

$$\hat{Q}_{Y_t}(\rho|\mathbf{X}) = \hat{Q}_{Y_t|\mathbf{X}_t}(\rho|\mathbf{X}) = \mathbf{X}_t'\boldsymbol{\beta}_\rho, for\ \rho \in (0,1)\ and\ t\ =\ 12\ or\ 15 \tag{4.4}$$

Then, we can obtain the simulated cumulative distribution function for each year, $\hat{F}_{Y_t|\mathbf{X}_t}(Y|\mathbf{X})$, by inverting the estimated quantile function for that year:

$$\hat{F}_{Y_t|\mathbf{X}_t}(y|\mathbf{X}) = \int_0^1 1(\hat{Q}_{Y_t|\mathbf{X}_t}(\rho|\mathbf{X}) \leq y)d\rho, for\ t\ =\ 12\ or\ 15 \tag{4.5}$$

The counterfactual distribution of house price in 2015 is obtained by integrating the estimated 2015 conditional distribution of house price with the 2012 distribution of covariates, as if the distribution of housing characteristics had remained as it was in 2012:

$$\hat{F}_{Y_{15}|\mathbf{X}_{12}}(y) = \int \hat{F}_{Y_{15}|\mathbf{X}_{15}}(y|\mathbf{X})d\hat{F}_{\mathbf{X}_{2012}}(\mathbf{X}) \tag{4.6}$$

where Chernozhukov et al. (2013) propose to estimate the 2012 distribution of covariates $F_{\mathbf{X}_{12}}(\mathbf{X})$ using the following empirical distribution function:

$$F_{\mathbf{X}_{12}}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} 1(\mathbf{X}_{12,i} \leq \mathbf{X}) \tag{4.7}$$

Change in the distribution of house prices observed between 2012 and 2015 can then be decomposed as follows:

$$\hat{F}_{Y_{15}}(y) - \hat{F}_{Y_{12}}(y) = \underbrace{[\hat{F}_{Y_{15}|\mathbf{X}_{15}}(y) - \hat{F}_{Y_{15}|\mathbf{X}_{12}}(y)]}_{\text{Composition}} + \underbrace{[\hat{F}_{Y_{15}|\mathbf{X}_{12}}(y) - \hat{F}_{Y_{12}|\mathbf{X}_{12}}(y)]}_{\text{Coefficient}} \tag{4.8}$$

The first term in the square brackets captures the "composition effect", the estimated change in the distribution of house prices induced by changes in the distributions of covariates across two years. The second term in the square brackets measures the "coefficient effect", the changes attributable to the structural variations in the relationship of house prices to housing characteristics, i.e., the change in the distribution of house prices conditional on a given distribution of characteristics.

The counterfactual distribution $\hat{F}_{Y_{15}|\mathbf{X}_{12}}(y)$ can be inverted to obtain the counterfactual quantile function:

$$\hat{Q}_{Y_{15}|\mathbf{X}_{12}}(\rho) = \hat{F}^{-1}_{Y_{15}|\mathbf{X}_{12}}(\rho) \tag{4.9}$$

We present the aggregate decomposition of house prices changes in quantiles as

$$\hat{Q}_{Y_{15}}(\rho) - \hat{Q}_{Y_{12}}(\rho) = \underbrace{[\hat{Q}_{Y_{15}|\mathbf{X}_{15}}(\rho) - \hat{Q}_{Y_{15}|\mathbf{X}_{12}}(\rho)]}_{\text{Composition}} + \underbrace{[\hat{Q}_{Y_{15}|\mathbf{X}_{12}}(\rho) - \hat{Q}_{Y_{12}|\mathbf{X}_{12}}(\rho)]}_{\text{Coefficient}} \tag{4.10}$$

Standard errors are calculated using the bootstrap method (500 iterations). The validity of this inference procedure is established by Chernozhukov et al. (2013).

4.3. Data

As the information technology developed in China, more and more people are inclined to search for information on the Internet. And the Internet has become a common platform to get information like housing characteristics and prices. We compiled a dataset of housing information for the period between 2012 to 2015 from a Chinese real estate website SouFun[3]. SouFun is one of the largest comprehensive housing website that provides real estate information of both newly-built and second-hand residential private housing units across 606 cities in China. We only collected housing data of newly-built units, as Wu et al. (2014) indicate, the newly-built units are a major source of all private housing units in terms of floor area at the national level. The database contains detailed housing characteristics such as transaction price, area, number of bedrooms, number of living rooms, address, and sold date etc.

For the period from 2012 to 2015 in Beijing, we collected 260,366 newly-built housing unit transactions in 8219 complexes from SouFun. We impose several restrictions on the data. First, we drop out duplicate records in terms of all characteristics and abandon the observations missing key housing characteristics reported in table 3.2. Then, we exclude extreme observations that are very high or low (out of four standard deviations) in terms of transaction prices, unit price per square meter, number of bedrooms, and number of living rooms. Finally, we drop out samples that are extremely large or small according to area and floor area ratio (out of eight standard deviations). For this analysis, we only use the housing transaction data in 2012 and 2015. We end up with 8511 transaction records in 2012 and

[3]The address of SouFun is http://www.fang.com/. The dataset was obtained automatically through a web scraper written by Python.

55850 records in 2015[4]. To keep balance in the number of record between two years, we randomly selected 8511 records without repetitions from the dataset of 2015[5].

The variables in this analysis are grouped into three categories: unit-level attributes which contain information that is unique to each housing unit, complex-level characteristics that are shared by housing units in the same complex, and geographic attributes relating to the location of the complex. For each transaction record, unit-level variables contains transaction price(in logarithm), number of bedrooms, number of dinning rooms, area, floor number, a villa dummy, and face directions of the unit(housing units facing north is used as the base group) ; complex-level variables include greening space ratio, floor area ratio, whether the complex has a sightseeing pool, and whether the complex a high-rise building; geographic attributes include whether the complex is near a subway station, whether the complex is near a park, and which ring road the complex is located, the base group is housing units located outside of the sixth ring road. Table 3.2 presents the definition of variables while table 4.1 shows descriptive statistics for both years. Throughout this study, we use natural logarithms of transaction prices as the dependent variable in regression analysis and objects for comparison in decomposition analysis.

There is substantial appreciation in house prices from 2012 to 2015. The average price of houses in 2012 is 340.925 ten thousands Yuan RMB while the corresponding number in 2015 is 514.864, a 51.3% appreciation in terms of nominal house price. We can notice that there are some variations on unit-level attributes from 2012 to 2015. The average area, number of bedrooms, and dinning rooms increased from 102.523, 2.126, and 1.34 to 123.94, 2.415 and 1.559, respectively. The increase of major unit-level attributes such as area may be one source of house prices appreciation but we are not clear to what extent they contribute

[4] Since we compiled data in January 2016, we notice that there are fewer records in previous years like 2011 and 2012 than recent years 2014 and 2015. One possible reason is that when all housing units in one complex are sold out, housing information about this complex is not listed in the newly-built housing sector of the website. This leads to more housing units that are constructed in recent years are compiled.

[5] We compare the summary statistics of key variables in the new sample and the original dataset and find that they are close to each other, then we argue that the sample is a good representative of the original dataset.

Table 4.1. Summary Statistics.

	2012				2015			
Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
1. Unit-level attributes								
price	340.925	413.743	15	12880	514.864	540.264	28	10000
lnprice	5.544	0.683	2.708	9.463	5.89	0.844	3.332	9.210
Bedrooms	2.126	0.974	0	21	2.415	1.107	0	21
DiningRooms	1.34	0.608	0	6	1.559	0.649	0	6
area	102.523	84.225	25	2600	123.94	84.914	6	1500
floor	14.099	8.560	0	101	15.342	9.645	0	63
Villa	0.042	0.202	0	1	0.104	0.305	0	1
North	0.041	0.198	0	1	0.04	0.196	0	1
West	0.04	0.196	0	1	0.031	0.173	0	1
East	0.049	0.216	0	1	0.043	0.204	0	1
South	0.159	0.365	0	1	0.165	0.371	0	1
Northwest	0.028	0.164	0	1	0.026	0.16	0	1
Northeast	0.035	0.183	0	1	0.029	0.167	0	1
Eastwest	0.037	0.188	0	1	0.023	0.149	0	1
Southwest	0.062	0.24	0	1	0.051	0.219	0	1
Southeast	0.072	0.259	0	1	0.07	0.256	0	1
Southnorth	0.478	0.5	0	1	0.521	0.5	0	1
2. Complex-level characteristics								
Green	0.344	0.075	0.004	0.8	0.354	0.078	0.1	0.85
FAR	2.525	1.271	0.1	10	2.503	1.197	0.08	10.7
WaterView	0.015	0.12	0	1	0.016	0.127	0	1
HighRise	0.693	0.461	0	1	0.768	0.422	0	1
3. Geographic attributes								
NearSubway	0.358	0.479	0	1	0.4	0.49	0	1
NearPark	0.029	0.168	0	1	0.052	0.221	0	1
ringRoad02	0.051	0.22	0	1	0.052	0.223	0	1
ringRoad23	0.155	0.362	0	1	0.095	0.293	0	1
ringRoad34	0.249	0.433	0	1	0.199	0.4	0	1
ringRoad45	0.245	0.43	0	1	0.239	0.426	0	1
ringRoad56	0.272	0.445	0	1	0.287	0.452	0	1
ringRoad600	0.028	0.165	0	1	0.128	0.334	0	1
N	8511				8511			

to the increase of real estate prices. There is also an increase in the average floor, indicating a trend of higher buildings in recent years. The portion of villa is only 4.02% in 2012 but increased to 10.4% in 2015, which means the quantities of villa constructed increased in the recent period. The three years saw no substantial changes on the face directions of housing unit, the housing units face both north and south("nan bei tong tou" in Chinese) dominate the real estate market, accounting for about half (0.478 in 2012 V.S. 0.521 in 2015) of total houses.

In terms of the complex-level characteristics, we also see no significant variation on floor-area ratio and whether the complex has a water view. Even though our observations are not completely consistent with the pattern that Brueckner et al. (2016) find in Beijing, an indication of continuation in FAR regulation of Beijing planners. There is a significant rise in the percentage of high-rise buildings, from 46.1% in 2012 to 76.8% in 2015, which is consistent with the increase of average floor.

We also control the geographic attributes that may affect house prices. For example, the proportions of complexes that are near a subway station or park experienced a major rise from 35.8% and 2.9% to 40% and 5.2%, respectively. As urbanization and income level advance in China, Zheng et al. (2009) show that urban residents become more and more willing to pay for local amenities. The rise of share of complexes near a subway station and park may reflect the developer's strategy to cater for the demand of high quality of living. Xu et al. (2015) also find evidence that people's willingness to pay for subway proximity increased after private driving had been restricted by the Beijing City government in October, 2008. It is also worth noting there is a trend that the newly-built residential real estate moves outside of the city: though the share of complexes that are within the second ring road, between the 4th and the 5th ring road, and between the 5th and the 6th ring road remained stable during the period, there is a substantial decline in the proportions of housing developments that are between the 2nd and the 3rd ring road and between the 3rd and the 4th ring road while a significant increase in the percentage of houses that are outside of the 6th ring road(from 2.8% in 2012 to 12.8% in 2015). This maybe because the land zoned for residential in the city core became scarce as the rapid pace of urbanization in Beijing in the past decade. Therefore, developers have to utilize land that is far away from the city center.

To show a clear picture of the distributional variation on house prices between 2012 and 2015, the density function for the logarithm of transaction prices in 2012 and 2015 are presented in 4.1. Transaction price distribution in 2015 shifted to the right but also contains a higher portion in the lower segment. The price distribution was more skewed in 2012, thicker on the middle-price segment than the right side of the distribution. Based on the result of two-sample Kolmogorov-Smirnov test, it is safe to reject the null hypothesis that the house prices(in logarithm form) in 2012 and 2015 come from the same distribution(p-value = 0.000)[6].

Figure 4.1. Kernel Density Estimates for Log of Sales Price.

The cumulative density function is shown in figure 4.2. The horizontal distance between the two densities decrease from the lowest percentiles to zero, then increase to the highest percentiles. For example, the 5th percentile of log house price is 4.605 in 2012, which is larger than the corresponding number in 2015, 4.372 - a difference of 0.232. The 10th

[6]We also check the two-sample Kolmogorov-Smirnov test on the unit house prices(house prices divided by the area), the null hypothesis that the unit house prices in 2012 and 2015 come from the same distribution is also rejected (p-value = 0.000).

percentile of the distribution takes a value of 4.771 in 2012, compared with 4.816 in 2015, a distance of 0.045 but the house price in 2015 is larger. In the higher percentiles like 50th and 90th, the differences rise to 0.437 and 0.480, respectively. The implication is that there was depreciation of house prices in lowest percentiles but the rate of depreciation declined to zero at the 9th percentile. After the 9th percentile, there is appreciation of house price and the rate of appreciation was larger at higher percentiles, which leaded to highest appreciation rate for high-priced houses.

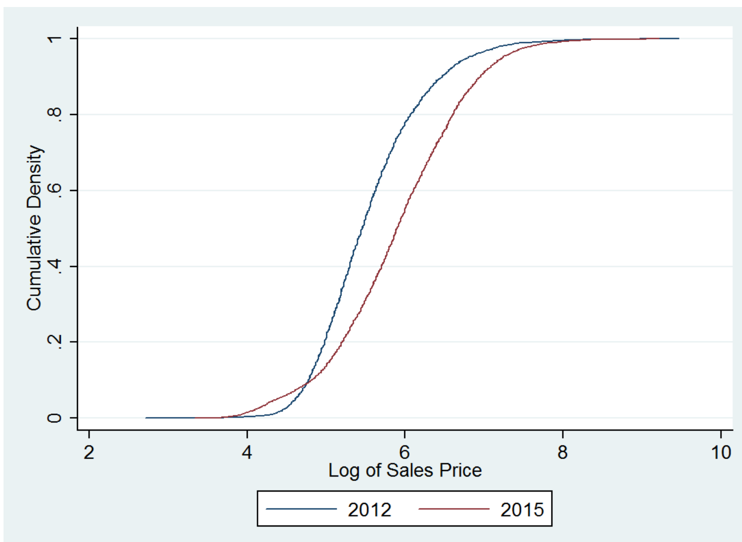

Figure 4.2. Estimated Cumulative Density Function for Log of Sales Price.

4.4. Results

4.4.1. Results of OLS Regression

We employ OLS regression method to examine the house price determination. The dependent variable is the logarithmic transaction price. Explanatory variables in this study include a comprehensive set of housing attributes defined in table 3.2 except monthly sold dummies. The coefficients of OLS estimation can be explained as the marginal effects of covariates on the conditional mean of log house prices. The OLS estimation results are presented in the second column and six column of table 4.2 for year 2012 and 2015, respectively.

Table 4.2. Regression Results via OLS and QR

	2012				2015			
	OLS	Q(0.25)	Q(0.50)	Q(0.75)	OLS	Q(0.25)	Q(0.50)	Q(0.75)
Bedrooms	0.150***	0.0994***	0.0933***	0.0649***	0.0719***	0.0668***	0.0407***	0.0312***
DiningRooms	0.163***	0.144***	0.109***	0.0902***	0.152***	0.163***	0.137***	0.129***
area	0.00319***	0.00512***	0.00653***	0.00729***	0.00552***	0.00605***	0.00703***	0.00726***
floor	0.0107***	0.0106***	0.00733***	0.00543***	0.000973*	0.00297***	0.000832	-0.000480
Villa	0.358***	0.191***	0.0878**	0.0487	-0.0778***	-0.0912***	-0.165***	-0.110***
West	0.0525*	0.0519*	0.0315	0.0577*	0.0471	0.0709*	0.0496	0.0711*
East	0.0980***	0.0553*	0.101***	0.118***	0.0721***	0.0947***	0.0761**	0.109***
South	0.127***	0.102***	0.0867***	0.123***	0.0644***	0.101***	0.0717***	0.0887***
Northwest	0.0657**	0.0267	0.0187	0.0755**	0.0760**	0.133***	0.0907***	0.0503*
Northeast	0.0962***	0.0899***	0.0653**	0.0961***	0.0576*	0.141***	0.0549	0.0365
Eastwest	0.189***	0.189***	0.143***	0.143***	0.207***	0.262***	0.192***	0.172***
Southwest	0.129***	0.120***	0.0936***	0.0956***	0.178***	0.166***	0.141***	0.143***
Southeast	0.148***	0.129***	0.100***	0.136***	0.142***	0.159***	0.130***	0.107***
Southnorth	0.269***	0.202***	0.166***	0.212***	0.214***	0.234***	0.183***	0.182***
Green	0.776***	0.296***	0.397***	0.540***	0.575***	0.576***	0.601***	0.420***
FAR	0.00689*	0.00512	0.00611	0.00475	0.00416	0.0133***	-0.00289	-0.0218***
WaterView	0.150***	0.00490	-0.0203	0.0517	0.215***	0.179***	0.165***	0.0717
HighRise	0.123***	0.113***	0.134***	0.104***	0.0770***	0.0939***	0.103***	0.109***
NearSubway	0.0140	0.0119	0.0217**	-0.00419	0.0371***	0.0517***	0.0198**	0.0251***
NearPark	0.0773***	0.141***	0.0662***	0.0768**	0.103***	0.148***	0.0799***	0.0717***
ringRoad02	1.027***	1.074***	1.008***	0.979***	1.662***	1.837***	1.736***	1.574***
ringRoad23	0.885***	0.919***	0.901***	0.880***	1.541***	1.712***	1.625***	1.459***
ringRoad34	0.883***	0.896***	0.856***	0.850***	1.482***	1.625***	1.532***	1.404***
ringRoad45	0.732***	0.710***	0.722***	0.736***	1.324***	1.524***	1.405***	1.205***
ringRoad56	0.396***	0.408***	0.385***	0.427***	0.964***	1.083***	1.016***	0.941***
Constant	3.252***	3.166***	3.364***	3.575***	3.217***	2.690***	3.112***	3.586***
N	8511	8511	8511	8511	8511	8511	8511	8511
adj./pseudo R^2	0.688	0.430	0.473	0.524	0.809	0.586	0.588	0.594

The standard errors for quantile regression are obtained through 500 bootstrap replications.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

housing units facing both south and north are the most popular style and are valued most in Beijing.

In addition, it appears that house price increase significantly as the greening rate of the complex rise in both years. The percentage effects of greening rate fell from 77.6% in

2012 to 57.5% in 2015. The positive correlation between floor-area ratio and house price is unexpected since higher floor-area ratio indicates more crowed living condition in the complex, however, the effect is pretty small in 2012 and even not significant in 2015. Houses that have a water view and are close to a park are worth relatively more in 2015 than in 2012, a price premium of 21.5% to 15% and of 10.3% to 7.73%, which indicates a rising willingness to pay for pleasing living environment (Zheng et al., 2009). As we can see from the descriptive statistics, the percentage of high-rise building increase from 2012 to 2015, unsurprisingly, housing units in a high-rise complex are worth more in 2012–the coefficients on high-rise suggest a premium of roughly 12.3% in 2012 and of 7.7% in 2015 on average. The results are somewhat consistent with the finding of estimated relative effect of floor, i.e., high-rise and high floor housing units are valued more in 2012 than in 2015.

There is a significant increase in percentage effect on whether a house is close to a subway station, 1.40% premium(not significant) in 2012 comparing with a larger and significant premium of 3.71% in 2015. Nowadays, as traffic congestion problem worsen in Beijing, there maybe a increasing demand for subway(Xu et al., 2015). A clear trend is found that housing units that are within the sixth ring road are significantly cheaper in 2012 than in 2015. For instance, the percentage effect of real estate that are within the second ring road are 102.7% in 2012 while increased significantly to 166.2% in 2015 compared with the base group that houses are outside the sixth ring road in each year. Also, coefficients of Ringroad56 suggest a price premium of only 39.6% in 2012 but of up to 96.4% in 2015.

In general, the housing characteristics we examined are shown to be important determinants of house prices in both years. Moreover, we see that the average price effects of housing attributes have changed between 2012 and 2015, which calls for a decomposition analysis to divide the increase in mean house prices into the effect caused compositional variations in housing characteristics and the effect due to changes in the underlying house price function.

4.4.2. Quantile Regression Results

The OLS estimation results provide a primary sense of determination of house prices. To give further information into the pattern of coefficients across the full distribution, we estimated 19 quantile regressions for quantiles ranging from 0.05 to 0.95 in increments of 0.05. We present coefficient estimates of the representative (0.25, 0.50, and 0.75) quantiles in table 4.2 and more details in tables A.2 and A.3. Further, selected scatter plots of the coefficient estimates by quantile are presented in figure 4.3 while the left are shown in figure A.1. Quantile estimates are graphed in solid lines with color red and blue to make a distinction between year 2012 and 2015. The associated confidence intervals are also included (dashed-dotted lines) only for quantile efficients. The OLS estimates are graphed in dashed lines for comparison.

The quantile regression results indicate that a mean regression provides limited information. Significant quantile effects are found for several housing attributes, i.e., the estimated price returns to these attributes vary across different quantiles. One more bedroom adds higher percentage house values at low quantiles than high quantiles in both years. The difference between mean and quantile estimates increases across the distribution while the gap is larger in 2012 than in 2015. For example, price premiums at quantile 0.75 are 6.49% in 2012 and 3.12% in 2015, a difference of 8.51% to the mean in 2012 while the gap is only 4.07% in 2015. One more living room exhibits a similar pattern but the price premium gap is not that large like bedroom. However, the gap of the estimated coefficients between 2012 and 2015 increases across the distribution. For instance, add one living room on the median rise house price by 10.9% in 2012 and 13.7% in 2015, a gap of 2.8%, while at quantile 0.75, the gap increases to 3.27% – 9.02% and 12.9% in 2012 and 2015, respectively. There is a significant upswing of coefficients for area from lower points of house price distribution to higher points, i.e., more areas increased house prices by more at higher quantiles, especially in 2012.

Quantile effects are mixed for greening rate. We can observe a significant rise of the coefficient in 2012 but in 2015 an almost reversed pattern exhibited. There is a swing to the mean coefficient at low and middle quantiles but a shift downward from the median to high quantiles. In terms of face directions of houses, the evolvement patterns of coefficients across different directions are almost the same. For example, the direction of south and north did have a statistically significant effect in the quantile regression in both years, but the gap between the mean coefficient and the quantile coefficients did not show a clear trend, i.e. the quantile coefficients sway around the mean coefficient. The positive effects of floor-area ratio on house prices are surprising but they are not significant and are close to zero across the distribution in 2012. In 2015, the floor-area ratio had a significant negative relationship with house prices from the median to the top end and illustrate a significant drop, suggesting that the transaction prices of higher quantiles are more sensitive to the change in the floor-area ratio than the lower part the house prices. The price premiums of whether a water view is available are not significant across quantiles in 2012, while a water view rose prices at low and middle quantiles but the rise decreased significantly at high quantiles in 2015, indicating that water view is relatively worth more at low quantiles. The reason maybe buyers of low-price houses care more about living conditions because high-quality living space is often not accessible to them in big cities like Beijing.

The marginal effect of near subway on house prices demonstrates substantial heterogeneity across quantiles, as shown in figure 4.3. There is a significant rise from 2012 to 2015 among the whole distribution especially at low quantiles. Take quantile 0.25 as an example, the price premium of near subway increased from 1.19% (insignificant) in 2012 to 5.17% (significant) in 2015, which means the prices of housing units close to a subway station are roughly 5% higher than houses that are not close to a subway station. As the population of residents and quantity of private cars in Beijing increased substantially during recent years(Xu et al., 2015), the traffic congestion is and will be a tough problem facing the city authority and also every residents. But the demand for subway is not the same for house buyers at different quantiles of prices. The demand is substantially larger for low quantiles and drops significantly from the lowest portion of house price distribution. Similarly, near a park added more house values at lower price quantiles in both years. In 2015, the marginal effect is 14.8% at the 25th quantile while only 7.17% at the 75th quantile. Quantile effects are evident for regional variables “ringRoad02” to “ringRoad56”, as a representative plot of “ringRoad02” illustrated in figure 4.3 and the rest graphs in figure A.1. In 2012, the magnitude of the impact decreased from 122.9% at the 10th percentile to 103.3% at the 40th percentile and then followed a pattern of swaying around the mean. While in 2015, the magnitude varied over the entire distribution, starting at a premium of 189.3% at the very left of the distribution (quantile 0.10), monotonously dropped downward 142.4% at the right tail (quantile 0.90). This means houses are valued relatively higher at lower quantiles than higher quantiles.

4.4.3. Mean Decomposition

As shown by the descriptive statistics in table 4.1, some housing characteristics, such as the number of bedrooms and area, have changed substantially between 2012 and 2015. We also see from the OLS regression in table 4.2 that the mean estimated coefficients of housing attributes differ in two periods. One question is that to what extent the change in house prices between 2012 and 2015 was due to the change in the housing related attributes, and how much was due to the variation on the hedonic pricing function coefficients. We use the Oaxaca-blinder decomposition method to separate the mean log house price gap $\hat{\triangle} Y = \bar{Y}_{2015} - \bar{Y}_{2012}$ into two parts: one part is the composition effect that are attributed to the variation on the housing characteristics $\hat{\triangle} Y^x = \hat{\boldsymbol{\beta}}_{15}(\bar{\mathbf{X}}'_{15} - \bar{\mathbf{X}}'_{12})$; the other part is the coefficient effect that are induced by the differences in the returns to housing characteristics $\hat{\triangle} Y^u = \bar{\mathbf{X}}'_{15}(\hat{\boldsymbol{\beta}}_{15} - \hat{\boldsymbol{\beta}}_{12})$. The second row of table 4.3 presents the mean decomposition results.

There was 51.3% appreciation of the average house price from 340.925 in 2012 to 514.864 in 2015, bringing on a log house price gap of 0.346 between 2012 and 2015. In

Figure 4.3. Coefficient Estimates by Quantile (Selected).

Note: 2012 values in red, 2015 values in blue; the horizontal dashed line in each figure is the OLS estimate; the solid line is the quantile estimate and associated confidence interval (dashed and dotted).

which change in housing characteristics contributed about 0.046, roughly 13.3% of mean price differentials, as shown in the Oaxaca-blinder decomposition results. The rest of mean price gap, 86.7% of mean price differential, 0.300, was induced by the variation on the coefficients of housing characteristics between 2012 and 2015. Therefore, the change in the estimate coefficients between 2012 and 2015 rather than the temporal evolvement of housing attributes played a dominant role in explaining the mean house price gap.

Table 4.3. Aggregate Decomposition Results.

Quantile	Total	Std.E	Composition	Std.E	Percentage	Coefficient	Std.E	Percentage
mean	0.346	0.012	0.046	0.010	13.3%	0.300	0.006	86.7%
.05	-0.160	0.022	-0.394	0.023	-244.2%	0.233	0.013	144.2%
.1	0.057	0.019	-0.235	0.018	-415.3%	0.292	0.010	515.3%
.15	0.176	0.015	-0.139	0.014	-79.3%	0.315	0.009	179.3%
.2	0.245	0.013	-0.082	0.012	-33.4%	0.327	0.008	133.4%
.25	0.291	0.013	-0.044	0.011	-15.0%	0.335	0.008	115.0%
.3	0.325	0.013	-0.014	0.010	-4.3%	0.339	0.008	104.3%
.35	0.352	0.013	0.010	0.010	3.0%	0.342	0.007	97.0%
.4	0.375	0.013	0.032	0.010	8.7%	0.342	0.008	91.3%
.45	0.395	0.013	0.053	0.010	13.4%	0.342	0.007	86.6%
.5	0.413	0.013	0.073	0.010	17.7%	0.340	0.008	82.3%
.55	0.431	0.014	0.093	0.010	21.5%	0.338	0.008	78.5%
.6	0.447	0.014	0.113	0.011	25.3%	0.334	0.008	74.7%
.65	0.464	0.015	0.135	0.011	29.1%	0.329	0.008	70.9%
.7	0.480	0.015	0.155	0.012	32.3%	0.325	0.008	67.7%
.75	0.499	0.016	0.179	0.013	35.8%	0.320	0.008	64.2%
.8	0.516	0.017	0.204	0.014	39.5%	0.312	0.008	60.5%
.85	0.530	0.019	0.227	0.016	42.8%	0.303	0.008	57.2%
.9	0.547	0.021	0.252	0.019	46.1%	0.295	0.010	53.9%
.95	0.583	0.029	0.311	0.028	53.4%	0.272	0.016	46.6%

We may examine the determinants of house price for possible explanations. On one hand, we can note that some important price determinants did increase from 2012 to 2015. For example, the mean value of area rose by 20.89% from 102.523 square meters in 2012 to 123.94 square meters in 2015. Similarly, the average number of bedroom and living rooms also increased by 13.59% and 16.34%, respectively. The rise of these attributes would contribute to increase of house price. On the other hand, we can also observe a trend that more recent developments were constructed from outside the city's central urban area into suburban and outer-suburban districts. Since geography attributes like the location of houses are also important determinants of house prices, especially in big cities like Beijing. The tendency of sub-urbanization would cause house prices to move downward. The contribution of determinants, like area, that move house prices upward was partly offset by some additional attributes, like the location variable ring road. These attributes in all lead the composition effect only slightly responsible for the mean price gap.

4.4.4. Quantile Decomposition

The mean decomposition results indicate that changes in housing characteristics did not have a crucial effect on the house price gap between 2012 and 2015. One natural question to ask is whether the effects of changes in housing characteristics are similar across the house price distribution? So in this section, we further develop insight into the house price distributions of 2012 and 2015 utilizing quantile decomposition methods. We decompose the logarithmic house price gaps $(\hat{Q}_{Y_{15}}(\rho) - \hat{Q}_{Y_{12}}(\rho))$ for a fixed quantile ρ into a composition effect: $\hat{Q}_{Y_{15}|\mathbf{X}_{15}}(\rho) - \hat{Q}_{Y_{15}|\mathbf{X}_{12}}(\rho)$, and a coefficient effect: $\hat{Q}_{Y_{15}|\mathbf{X}_{12}}(\rho) - \hat{Q}_{Y_{12}|\mathbf{X}_{12}}(\rho)$. Table 4.3 presents the aggregate decomposition results for selected quantiles of house price distribution from 0.05 to 0.95 in increments of 0.05. Figure 4.4 provides an illustration graph of both the quantile decomposition and the mean decomposition results for comparison.

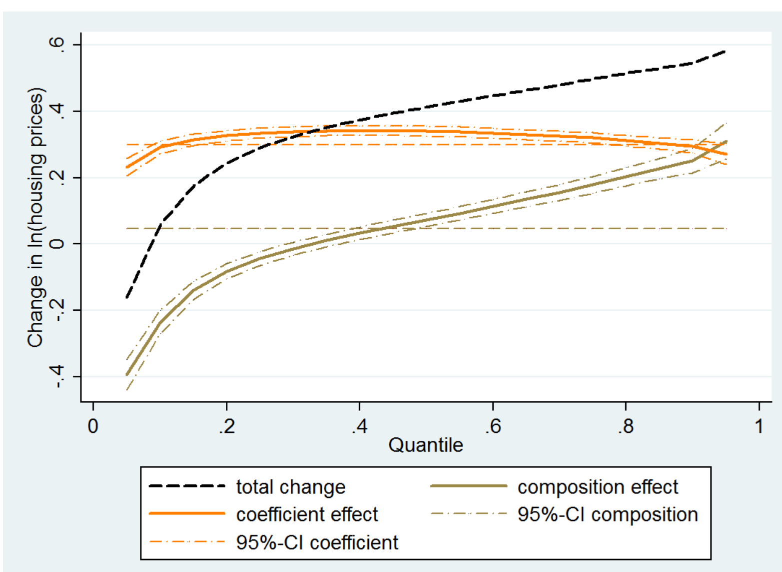

Figure 4.4. Aggregated Quantile Decomposition Results with 95% Confidence Intervals.

The dashed line in the first plot of figure 4.4 is the house price gaps for different quantiles. The quantile price gap starts from roughly -16% at the 5th quantile of the distribution, then increase monotonously up to more than 58% at the 95th quantile. The solid orange line in figure 4.4 represents the coefficient effect of the aggregate decomposition. The quantile decomposition results in table 4.3 and figure 4.4 show that the coefficient effects are pretty stable across the quantiles, with a minimum effect of roughly 23.3% at the left tail of the distribution and rise up slowly to about 34% at the median, then move down slightly to about 27% at the right tail. In contrast, the solid brown line in figure 4.4 shows the composition effect experience a upsurge from 39% at 5th quantile of the distribution to about -8% at the second percentile, then increase almost in proportion to 31.1% at the right end of 95th quantile. The 95% confidence interval of the composition and coefficient effect are also shown in the second panel of figure 4.4, we can see the confidence interval are close to the estimates and the two confidence interval are far away from each other, which is an indicator of accuracy of the statistics produced by quantile decomposition.

As shown in the results of quantile decomposition, even the house price gap between year 2012 and 2015 is positive from the 10th quantile to the upper end of the distribution, the composition effect is negative under the 35th quantile, which means that housing characteristics are relatively better in 2012 than in 2015, resulting into a negative effect to the total change of house price. For example, at the first decile, the total difference of price is 0.057, of which -0.235 was induced by the variation of housing attributes between 2012 and 2015, while 0.292 was caused by altered implicit prices of housing attributes from 2012 to 2015. However, from the 35th quantile, a positive composition effect can be found, which indicates that housing characteristics in 2015 began to be relatively more valuable than the housing characteristics in 2015. At the 45th quantile, 13.4% of the house price gap is explained by the composition effect, which starts to exceed the mean level of 13.3%, as shown in the Oaxaca-blinder mean decomposition. While at the 45th quantile, the primary source of price change comes still the changes in coefficients, contributes to 86.6% of the total gap.

The upward pattern of the composition effect over the price distribution has been accompanied by a slight decrease in the coefficient effect, which leads to a monotonic increase in the percentage of composition effect and a decrease in the proportion of coefficient effect. Finally, at the 95th quantile, the composition effect contributed 0.311 to a total price of 0.583, a percentage of 53.4%, which indicates the relative contribution of the composition effect exceeds the coefficient effect.

The composition effect captures the changes related to size, quality, and location of the housing units. Previous research of McMillen (2008) and Nicodemo and Raya (2012) find that nearly all of the house price gap between two periods is explained by the coefficient effect over the whole distribution while the change in the housing characteristics have a limited effect on the price gap. Unlike the mature housing market in developed countries where transaction data primary comes from the secondary housing market, the housing market in China is nascent where a large proportion of transactions are newly-built housing units. Moreover, the housing market in Beijing experiences relatively great changes over a short period. In this sense, since we use transaction data coming from the newly-built developments, the decomposition results may be different from the results found in housing studies of developed countries like American and Spain. Even though the results of quantile decomposition shows that the coefficient effect plays a dominant rule to the house price differential than the composition effect at the lower quantiles of the distribution, the magnitude of effect is still larger than what McMillen (2008) and Nicodemo and Raya (2012) find. From the median quantiles to the upper end of the distribution, the magnitude of the composition effect shows that changes in the housing characteristics become a main source of price gap and even play an equivalent role with the coefficient effect in terms of the contribution to the price gap between 2012 and 2015. The negative and significant composition effect indicates even house prices increased at lower quantiles, the living condition in 2015 was worse for low-price home buyers regrading the housing attributes. While for the high-price home consumers, the living condition are even better in 2015 than in 2012, but at the expense of higher payment in 2015, the relatively higher payment came from both the improvement of housing characteristics and changes in the underlying hedonic price functions.

4.5. Conclusions

In this chapter, we analyze the changes in the distribution of house prices over time using regression and decomposition methods. Utilizing a comprehensive dataset about real estate transactions in 2012 and 2015, we show that house price increased by roughly 51% on average and the price gap varies with positions in the distribution over time. More low-price and high-price home sales lead the distribution of house price in 2015 thicker at the left and right tail. OLS and quantile regression results suggest that housing characteristics are important determinants of house price in both 2012 and 2015. Moreover, substantial heterogeneity and quantile effects are illustrated in the regression results, indicating that housing characteristics are valued differently between 2012 and 2015 and at different locations of the price distribution.

We further employ mean decomposition method developed by Oaxaca (1973) and (Blinder, 1973) and conditional quantile decomposition approach refined by Chernozhukov et al. (2013) to disentangle temporal price differential on average and at selected parts of the distribution into two components: one composition effect driven by changes in housing characteristics from 2012 to 2015, and one coefficient effect attributed to varying returns to housing characteristics in the underlying price functions. The results of Oaxaca-Blinder mean decomposition suggest the proportion of the price gap due to altered returns over time is greater than the proportion caused by changes in housing characteristics. Specifically, only 13% of price gaps are attributed to the composition effect. Findings in secondary housing market (McMillen, 2008; Nicodemo and Raya, 2012; Thomschke, 2015) suggest that changing returns over time are responsible for almost all price difference over the distribution, results of quantile decomposition in this study present substantial heterogeneity regarding the contributions of composition effect and coefficient effect to distributional price gaps. In this sense, our results are consistent with findings by (Qin et al., 2016) in decomposition analysis of the changes in the distribution of land prices in China. The composition effect rises monotonously from negative values at lower quantile to positive at higher quantiles while the coefficient effect remains positive and stable, suggesting that altered returns to housing characteristics contributed positively to house price over the full distribution though house price were lower in 2015 at the left tail of distribution. In addition, the signs and magnitude of the composition effect indicate that low-price house buyers paid more in 2015 even to worse housing characteristics, and high-price real estate consumers bought more favorable houses but at the cost of relatively higher expense.

5. CONCLUSIONS AND DISCUSSION

5.1. Conclusions

House price was, is, and will continuously be a hotting topic in China since it is related to almost every Chinese. Growing urbanization and rising income provoke people a "Chinese Dream" of owning a large housing unit in a big city. But skyrocketing house prices in China, especially in first-tier cities, have gained mass concerns, and prevent people from stepping into big cities like Beijing, Shanghai, etc. Recent rapid development of information technology enable us to collect sequential transaction data of new homes in Beijing on a leading Chinese real estate website. Utilizing by far the most comprehensive public data for newly-built residential housing units, this study provides insight into the evolution of Beijing's house price distributions and shed light on understanding changes in the distribution of house prices over time.

In the first part of this thesis, we employ a quantile regression approach on hedonic price model to construct house price indices across conditional distribution of house price. Important variables investigated contain unit-level housing characteristics such as number of bedrooms, number of living rooms, square meters of living area, face direction of the house, etc.; housing complex-level characteristics such as green space rate, floor area ratio, etc.; and geographic attributes such as close to subway, park, etc. Two main findings are shown in the first part. First, we show substantial appreciation of house price–17.6% to 27.1% for different quantiles–from January 2013 to February 2014, where the house prices peaked for all quantiles during our research period. Following that there is a sudden decline for the full distribution of house price from February 2014 to August 2014. After that house prices enter into a comparing slower appreciation process until December 2015, where the house price was still slightly lower than the peak reached in February 2014. Second, even though we show that similar temporal patterns are shared by house price indices across quantiles, substantial variation on the appreciation rate of house prices among different quantiles of house price distributions are presented. The appreciate rate is significantly higher for low-priced housing units than high-priced homes across almost all periods from January 2013 to December 2015. Moreover, examination of home buyers' marginal willingness to pay for a housing characteristic indicate that lower-priced home buyers value most housing attributes differently from higher-priced home buyers.

In the second part, observing that roughly 51% of appreciation on average and different distribution of house price between 2012 and 2015, we further analyze temporal changes in the distribution of house prices using regression (OLS and quantile) and decomposition methods (mean by Oaxaca (1973) and (Blinder, 1973); and conditional quantile by Chernozhukov et al. (2013)). We show housing characteristics have great power in explaining house price in both 2012 and 2015. In addition, regression results suggest evident quantile effects on housing attributes and substantial heterogeneity on returns to characteristics between 2012 and 2015, indicating that housing characteristics are valued differently over time and across the price distribution. Moreover, we decompose temporal price gaps on average and on selected quantile of the distribution into two part: a composition effect attributed to varying housing characteristics from 2012 to 2015, and a coefficient effect caused by changing returns to housing characteristics in the underlying price functions. Mean decomposition results suggest that only 13% of price gaps are attributed to changing housing characteristics while the most 87% of price gaps are the contribution of altered returns to house characteristics over time. Quantile decomposition results show substantial heterogeneity in terms of the contributions of the composition effect and the coefficient effect on distributional price gaps. The composition effect rises monotonously from negative at the left tail of distribution to positive from middle quantiles while the coefficient effect remains positive and stable, indicating that the coefficient effect rose house price over the full distribution though house price were lower in 2015 at lower quantiles. Finally, the implication of the signs and the magnitude of composition effect is that low-priced house buyers bought houses with worse housing characteristics in 2015 even with more payments compared with in 2012, and high-price home consumers obtained houses with more favorable housing attributes but at the expense of relatively higher payments.

5.2. Limitations and Plan for Future Research

Limited by time, data availability, and most importantly my knowledge in this area, this study only provides analysis of house prices in Beijing for a relatively short time period. The data set of house transaction only contains newly-built housing units. Future research can be extended at least to the second housing market in Beijing and also many other Chinese cities for comparison analysis. Moreover, even decomposition analysis at the aggregate level provides information to understand what caused the variation of distribution of house prices over time, detailed decomposition would provide more complementary results to assess contributions for a certain housing characteristic or a group of attributes. At last, recent development in the literature has witnessed a wide arrange of decomposition methods suggested by many scholars. Other decomposition methods may be employed in the future for a robustness check and also comparison of methods in an applied perspective.

REFERENCES

Alan S Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

Jan K Brueckner, Shihe Fu, Yizhen Gu, and Junfu Zhang. Measuring the stringency of land-use regulation: The case of china's building-height limits. Technical report, CESifo Group Munich, 2016.

Moshe Buchinsky. Changes in the us wage structure 1963-1987: Application of quantile regression. *Econometrica: Journal of the Econometric Society*, pages 405–458, 1994.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

John M Clapp and Carmelo Giaccotto. Estimating price indices for residential property: a comparison of repeat sales and assessed value methods. *Journal of the American Statistical Association*, 87(418):300–306, 1992.

Deborah A Cobb-Clark and Mathias G Sinning. Neighborhood diversity and the appreciation of native-and immigrant-owned homes. *Regional Science and Urban Economics*, 41(3): 214–226, 2011.

Peter F Colwell and Gene Dilmore. Who was first? an examination of an early hedonic study. *Land Economics*, pages 620–626, 1999.

N Edward Coulson and Daniel P McMillen. The dynamics of intraurban quantile house price indexes. *Urban Studies*, 44(8):1517–1537, 2007.

Andrew Court. Hedonic price indexes with automobile examples. *The dynamics of the automobile demand*, 1939.

Yongheng Deng, Randall Morck, Jing Wu, and Bernard Yeung. Monetary and fiscal stimuli, ownership structure, and china's housing market. Technical report, National Bureau of Economic Research, 2011.

Yongheng Deng, Daniel P McMillen, and Tien Foo Sing. Private residential price indices in singapore: a matching approach. *Regional Science and Urban Economics*, 42(3):485–494, 2012.

John DiNardo, Nicole M Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044, 1996.

Hanming Fang, Quanlin Gu, Wei Xiong, and Li-An Zhou. Demystifying the chinese housing boom. Technical report, National Bureau of Economic Research, 2015.

Michael Farmer and Clifford Lipscomb. Using quantile regression in hedonic analysis to reveal submarket competition. *Journal of Real Estate Research*, 32(4):435–460, 2010.

Eric Fesselmeyer, Kien T Le, and Kiat Ying Seah. Changes in the white–black house value distribution gap from 1997 to 2005. *Regional Science and Urban Economics*, 43(1):132–141, 2013.

Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition methods in economics. *Handbook of labor economics*, 4:1–102, 2011.

Allen C Goodman. Andrew court and the invention of hedonic price analysis. *Journal of urban economics*, 44(2):291–298, 1998.

Amanda Gosling, Stephen Machin, and Costas Meghir. The changing distribution of male wages in the uk. *The Review of Economic Studies*, 67(4):635–666, 2000.

Zvi Griliches. The demand for fertilizer: An economic interpretation of a technical change. *Journal of Farm Economics*, 40(3):591–606, 1958.

Xiaoyang Guo, Siqi Zheng, David Geltner, and Hongyu Liu. A new approach for constructing home price indices: The pseudo repeat sales model and its application in china. *Journal of Housing Economics*, 25:20–38, 2014.

George Casper Haas. A statistical analysis of farm sales in blue earth county, minnesota, as a basis for farm land appraisal. Technical report, University of Minnesota, Department of Applied Economics, 1922.

John P Harding, Stuart S Rosenthal, and CF Sirmans. Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model. *Journal of urban Economics*, 61(2):193–217, 2007.

Robert J Hill. Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, 27(5):879–914, 2013.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

Kelvin J Lancaster. A new approach to consumer theory. *The journal of political economy*, pages 132–157, 1966.

Wen-Chi Liao and Xizhu Wang. Hedonic house prices and spatial quantile regression. *Journal of Housing Economics*, 21(1):16–27, 2012.

Justin Yifu Lin. China and the global economy 1. *China Economic Journal*, 4(1):1–14, 2011.

Yinqiu Lu and Tao Sun. Local government financing platforms in china: A fortune or misfortune? Technical report, International Monetary Fund, 2013.

José AF Machado and José Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465, 2005a.

José AF Machado and José Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465, 2005b.

Stephen Mak, Lennon Choy, and Winky Ho. Quantile regression estimates of hong kong real estate prices. *Urban Studies*, 2010.

Daniel P McMillen. Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients? *Journal of Urban Economics*, 64(3):573–589, 2008.

Blaise Melly. Decomposition of differences in distribution using quantile regression. *Labour economics*, 12(4):577–590, 2005.

Blaise Melly. Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics*, 68(4):543–572, 2006.

Catia Nicodemo and Josep Maria Raya. Change in the distribution of house prices across spanish cities. *Regional Science and Urban Economics*, 42(4):739–748, 2012.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709, 1973.

Yu Qin, Hongjia Zhu, and Rong Zhu. Changes in the distribution of land prices in urban china during 2007–2012. *Regional Science and Urban Economics*, 57(1):77–90, 2016.

Sherwin Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, pages 34–55, 1974.

Stacy Sirmans, David Macpherson, and Emily Zietz. The composition of hedonic pricing models. *Journal of real estate literature*, 13(1):1–44, 2005.

Lorenz Thomschke. Changes in the distribution of rental prices in berlin. *Regional Science and Urban Economics*, 51:88–100, 2015.

Henry A Wallace. Comparative farm-land values in iowa. *The Journal of Land & Public Utility Economics*, 2(4):385–392, 1926.

Nancy E Wallace and Richard A Meese. The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics*, 14(1-2):51–73, 1997.

Ferdinand T Wang and Peter M Zorn. Estimating house price growth with repeat sales data: what's the aim of the game? *Journal of Housing Economics*, 6(2):93–118, 1997.

Jing Wu, Joseph Gyourko, and Yongheng Deng. Evaluating conditions in major chinese housing markets. *Regional Science and Urban Economics*, 42(3):531–543, 2012.

Jing Wu, Yongheng Deng, and Hongyu Liu. House price index construction in the nascent housing market: the case of china. *The Journal of Real Estate Finance and Economics*, 48(3):522–545, 2014.

Jing Wu, Joseph Gyourko, and Yongheng Deng. Evaluating the risk of chinese housing markets: what we know and what we need to know. Technical report, National Bureau of Economic Research, 2015.

Yangfei Xu, Qinghua Zhang, and Siqi Zheng. The rising demand for subway after private driving restriction: Evidence from beijing's housing market. *Regional Science and Urban Economics*, 54:28–37, 2015.

Lei Zhang and Tammy Leonard. Neighborhood impact of foreclosure: A quantile regression approach. *Regional Science and Urban Economics*, 48:133–143, 2014.

Siqi Zheng, Yuming Fu, and Hongyu Liu. Demand for urban quality of living in china: evolution in compensating land-rent and wage-rate differentials. *The Journal of Real Estate Finance and Economics*, 38(3):194–213, 2009.

Joachim Zietz, Emily Norman Zietz, and G Stacy Sirmans. Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics*, 37(4): 317–333, 2008.

APPENDIX

Table A.1. Summary Statistics of The Left Monthly Sold Dummies.

Variable	Mean	Std. Dev.	Min.	Max.
4.Month Dummies.				
Month 6	0.033	0.179	0	1
Month 7	0.038	0.192	0	1
Month 8	0.042	0.2	0	1
Month 9	0.045	0.206	0	1
Month 10	0.043	0.203	0	1
Month 11	0.042	0.201	0	1
Month 12	0.041	0.199	0	1
Month 13	0.033	0.178	0	1
Month 14	0.036	0.186	0	1
Month 15	0.044	0.206	0	1
Month 16	0.039	0.193	0	1
Month 17	0.032	0.177	0	1
Month 18	0.027	0.163	0	1
Month 19	0.029	0.167	0	1
Month 20	0.027	0.161	0	1
Month 21	0.022	0.146	0	1
Month 22	0.025	0.157	0	1
Month 23	0.018	0.132	0	1
Month 24	0.016	0.127	0	1
Month 25	0.02	0.141	0	1
Month 26	0.012	0.109	0	1
Month 27	0.022	0.147	0	1
Month 28	0.03	0.172	0	1
Month 29	0.028	0.165	0	1
Month 30	0.027	0.163	0	1
Month 31	0.029	0.168	0	1
Month 32	0.028	0.165	0	1
Month 33	0.027	0.162	0	1
Month 34	0.023	0.151	0	1
Month 35	0.025	0.156	0	1
Month 36	0.02	0.139	0	1
N		190580		

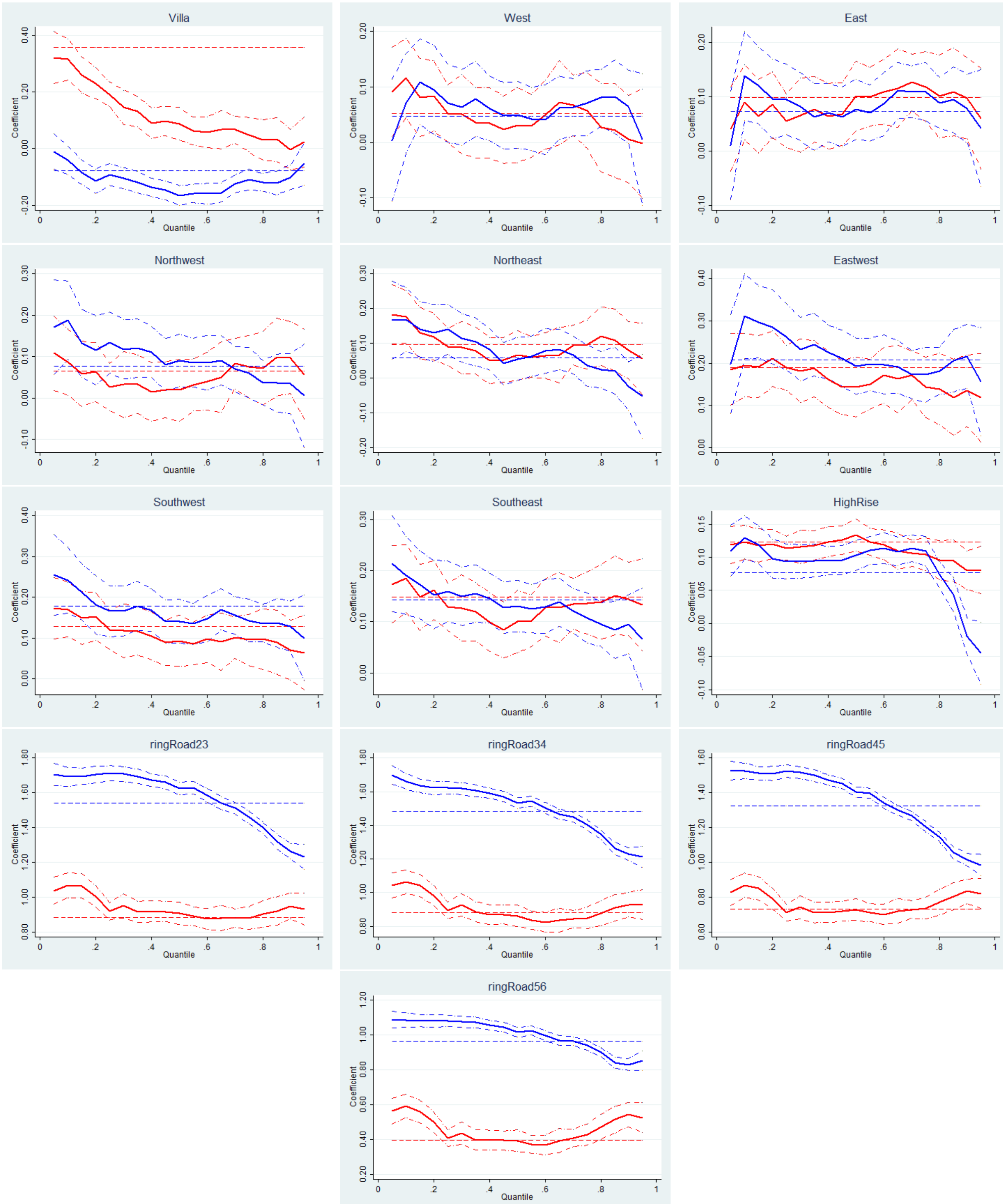

Figure A.1. Coefficient Estimates by Quantile (Remained).

Note: 2012 values in red, 2015 values in blue; the horizontal dashed line in each figure is the OLS estimate; the solid line is the quantile estimate and associated confidence interval (dashed and dotted).

Table A.2. Regression Results of 2012 via OLS and QR

	OLS	Q(0.10)	Q(0.20)	Q(0.30)	Q(0.40)	Q(0.50)	Q(0.60)	Q(0.70)	Q(0.80)	Q(0.90)
Bedrooms	0.150***	0.147***	0.113***	0.0957***	0.0946***	0.0933***	0.0850***	0.0659***	0.0499***	0.0421***
	(23.00)	(5.96)	(5.05)	(5.49)	(5.92)	(7.14)	(7.73)	(6.53)	(4.64)	(3.64)
DiningRooms	0.163***	0.175***	0.159***	0.128***	0.124***	0.109***	0.0932***	0.0940***	0.0731***	0.0522***
	(17.39)	(8.11)	(8.98)	(8.63)	(8.38)	(7.58)	(6.66)	(6.91)	(5.75)	(3.38)
area	0.0032***	0.0032***	0.0045***	0.0056***	0.0060***	0.0065***	0.0068***	0.0072***	0.0077***	0.0079***
	(44.65)	(5.53)	(7.72)	(12.42)	(15.63)	(21.76)	(24.98)	(29.23)	(28.01)	(30.91)
floor	0.0107***	0.0113***	0.0109***	0.0103***	0.0093***	0.0073***	0.0077***	0.0063***	0.0047***	0.0037***
	(15.65)	(12.09)	(13.87)	(12.12)	(10.92)	(8.35)	(9.65)	(8.66)	(5.69)	(3.58)
Villa	0.358***	0.314***	0.230***	0.146***	0.0899**	0.0878**	0.0574	0.0671**	0.0302	-0.0030
	(14.28)	(5.30)	(4.13)	(3.08)	(2.11)	(2.15)	(1.52)	(2.16)	(0.92)	(-0.08)
West	0.0525*	0.116***	0.0841**	0.0512*	0.0351	0.0315	0.0503	0.0672**	0.0267	0.0058
	(1.80)	(3.19)	(2.37)	(1.71)	(0.96)	(1.02)	(1.49)	(2.41)	(0.75)	(0.12)
East	0.0980***	0.0894**	0.0851**	0.0655*	0.0640*	0.101***	0.109***	0.126***	0.100***	0.0971**
	(3.54)	(2.03)	(2.44)	(1.88)	(1.76)	(3.16)	(4.25)	(4.80)	(2.99)	(2.50)
South	0.127***	0.174***	0.127***	0.0954***	0.0888***	0.0867***	0.101***	0.116***	0.119***	0.0937***
	(5.48)	(5.39)	(4.22)	(3.83)	(3.03)	(3.54)	(4.44)	(4.73)	(3.97)	(2.59)
Northwest	0.0657**	0.0865**	0.0626*	0.0324	0.0142	0.0187	0.0403	0.0828**	0.0717*	0.0974**
	(2.03)	(2.05)	(1.80)	(0.89)	(0.37)	(0.57)	(1.20)	(2.02)	(1.67)	(2.29)
Northeast	0.0962***	0.176***	0.118***	0.0877***	0.0502	0.0653**	0.0649**	0.0950***	0.120***	0.0803*
	(3.15)	(4.47)	(3.31)	(2.66)	(1.42)	(2.26)	(2.30)	(2.89)	(2.92)	(1.86)
Eastwest	0.189***	0.195***	0.210***	0.182***	0.161***	0.143***	0.171***	0.171***	0.138***	0.134**
	(6.19)	(4.73)	(5.61)	(5.66)	(4.98)	(4.32)	(5.95)	(5.92)	(3.91)	(2.57)
Southwest	0.129***	0.170***	0.153***	0.118***	0.106***	0.0936***	0.0984***	0.102***	0.0977***	0.0710
	(4.82)	(4.29)	(4.43)	(4.04)	(3.38)	(3.42)	(4.27)	(3.76)	(2.89)	(1.60)
Southeast	0.148***	0.185***	0.162***	0.126***	0.0994***	0.100***	0.128***	0.135***	0.138***	0.145***
	(5.63)	(5.08)	(5.10)	(4.63)	(3.29)	(3.53)	(4.89)	(5.15)	(3.93)	(3.33)
Southnorth	0.269***	0.248***	0.223***	0.203***	0.179***	0.166***	0.201***	0.220***	0.211***	0.215***
	(11.63)	(7.18)	(6.88)	(8.00)	(6.18)	(6.37)	(8.62)	(8.80)	(6.64)	(5.17)
Green	0.776***	0.318***	0.322***	0.373***	0.437***	0.397***	0.451***	0.486***	0.587***	0.717***
	(12.91)	(3.41)	(3.82)	(4.21)	(5.23)	(5.11)	(5.45)	(6.15)	(7.24)	(8.10)
FAR	0.0069*	0.0009	0.0010	0.0059	0.0063	0.0061	0.0023	0.0076	0.0069	0.0053
	(1.85)	(0.19)	(0.26)	(1.62)	(1.35)	(1.26)	(0.48)	(1.60)	(1.18)	(0.87)
WaterView	0.150***	-0.0325	0.0135	-0.0008	0.0073	-0.0203	0.0377	0.0497	0.0559	0.141*
	(4.17)	(-0.47)	(0.22)	(-0.02)	(0.12)	(-0.37)	(0.59)	(0.96)	(1.07)	(1.83)
HighRise	0.123***	0.123***	0.120***	0.116***	0.124***	0.134***	0.119***	0.106***	0.0955***	0.0804***
	(11.85)	(8.95)	(9.68)	(9.72)	(10.50)	(10.00)	(8.44)	(8.45)	(6.04)	(4.95)
NearSubway	0.0140	0.0184*	0.0082	0.0154	0.0209**	0.0217**	0.0135	0.0066	-0.0058	-0.0075
	(1.58)	(1.71)	(0.83)	(1.43)	(2.21)	(2.01)	(1.38)	(0.68)	(-0.53)	(-0.51)
NearPark	0.0773***	0.0857	0.122***	0.124***	0.0952***	0.0662***	0.0485	0.0487	0.0554*	-0.0255
	(3.02)	(1.63)	(2.90)	(4.74)	(3.62)	(3.29)	(1.49)	(1.39)	(1.81)	(-0.71)
ringRoad02	1.027***	1.229***	1.166***	1.088***	1.033***	1.008***	0.970***	0.975***	1.006***	1.029***
	(32.34)	(27.47)	(15.51)	(25.02)	(22.54)	(19.35)	(24.49)	(27.59)	(32.79)	(18.07)
ringRoad23	0.885***	1.069***	1.004***	0.950***	0.920***	0.901***	0.875***	0.881***	0.904***	0.949***
	(31.65)	(25.05)	(13.97)	(22.40)	(21.20)	(17.86)	(23.20)	(27.30)	(32.31)	(18.32)
ringRoad34	0.883***	1.065***	0.983***	0.929***	0.870***	0.856***	0.825***	0.843***	0.878***	0.929***
	(32.74)	(25.14)	(13.99)	(23.02)	(21.49)	(17.41)	(21.96)	(27.08)	(31.69)	(18.42)
ringRoad45	0.732***	0.869***	0.792***	0.741***	0.713***	0.722***	0.701***	0.729***	0.772***	0.835***
	(27.27)	(20.74)	(11.24)	(19.02)	(17.54)	(14.63)	(18.36)	(23.47)	(28.06)	(16.65)
ringRoad56	0.396***	0.592***	0.499***	0.436***	0.396***	0.385***	0.369***	0.408***	0.473***	0.542***
	(14.98)	(14.23)	(7.03)	(10.86)	(9.82)	(7.85)	(9.82)	(12.77)	(17.55)	(10.96)
Constant	3.252***	2.791***	3.003***	3.144***	3.248***	3.364***	3.458***	3.527***	3.610***	3.728***
	(81.74)	(41.05)	(36.04)	(54.95)	(56.75)	(56.29)	(69.22)	(78.27)	(80.70)	(49.08)
N	8511	8511	8511	8511	8511	8511	8511	8511	8511	8511
adj./pseudo R^2	0.688	0.396	0.420	0.438	0.455	0.473	0.492	0.513	0.537	0.567

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The standard errors for quantile regression are obtained through 500 bootstrap replications.

Table A.3. Regression Results of 2015 via OLS and QR

	OLS	Q(0.10)	Q(0.20)	Q(0.30)	Q(0.40)	Q(0.50)	Q(0.60)	Q(0.70)	Q(0.80)	Q(0.90)
Bedrooms	0.0719***	0.0709***	0.0668***	0.0617***	0.0501***	0.0407***	0.0341***	0.0311***	0.0329***	0.0203**
	(12.24)	(4.19)	(4.28)	(4.70)	(4.34)	(3.88)	(3.78)	(3.41)	(3.33)	(2.49)
DiningRooms	0.152***	0.149***	0.161***	0.154***	0.132***	0.137***	0.128***	0.131***	0.123***	0.120***
	(17.78)	(8.45)	(10.67)	(11.77)	(11.02)	(12.50)	(11.40)	(11.67)	(12.14)	(9.23)
area	0.0056***	0.0055***	0.0060***	0.0062***	0.0067***	0.0070***	0.0072***	0.0073***	0.0072***	0.0073***
	(69.97)	(24.78)	(23.47)	(28.19)	(28.76)	(37.26)	(40.96)	(40.44)	(40.45)	(37.19)
floor	0.0010*	0.0059***	0.0034***	0.0030***	0.0019***	0.0008	0.0004	-0.0002	-0.0010*	-0.0003
	(1.85)	(6.17)	(4.12)	(4.84)	(2.96)	(1.48)	(0.67)	(-0.39)	(-1.74)	(-0.44)
Villa	-0.0778***	-0.0415	-0.116***	-0.105***	-0.137***	-0.165***	-0.160***	-0.124***	-0.119***	-0.101***
	(-4.53)	(-1.43)	(-3.93)	(-3.98)	(-5.03)	(-5.59)	(-4.93)	(-3.96)	(-5.92)	(-3.24)
West	0.0471	0.0708	0.0948***	0.0633	0.0617*	0.0496	0.0412	0.0637*	0.0807**	0.0645*
	(1.55)	(1.13)	(2.97)	(1.61)	(1.67)	(1.60)	(1.35)	(1.67)	(2.37)	(1.80)
East	0.0721***	0.138**	0.0957***	0.0815**	0.0702*	0.0761**	0.0862***	0.109***	0.0882***	0.0782**
	(2.60)	(2.55)	(3.04)	(2.48)	(1.82)	(2.53)	(3.04)	(3.49)	(3.29)	(2.05)
South	0.0644***	0.110***	0.0894***	0.0928***	0.0794**	0.0717***	0.0924***	0.0978***	0.0789***	0.0796**
	(2.88)	(2.72)	(3.07)	(2.97)	(2.54)	(2.77)	(3.93)	(3.81)	(2.95)	(2.49)
Northwest	0.0760**	0.189***	0.115***	0.118***	0.111***	0.0907***	0.0854***	0.0698**	0.0369	0.0350
	(2.39)	(4.38)	(2.75)	(3.08)	(3.26)	(2.92)	(3.21)	(2.46)	(1.19)	(0.90)
Northeast	0.0576*	0.169***	0.130***	0.115***	0.0824**	0.0549	0.0781**	0.0664**	0.0240	-0.0231
	(1.86)	(3.59)	(3.31)	(3.38)	(2.53)	(1.60)	(2.48)	(2.31)	(0.81)	(-0.59)
Eastwest	0.207***	0.311***	0.286***	0.233***	0.226***	0.192***	0.197***	0.173***	0.181***	0.216***
	(6.19)	(4.54)	(7.41)	(5.47)	(5.28)	(5.55)	(6.63)	(4.82)	(4.24)	(4.32)
Southwest	0.178***	0.242***	0.182***	0.166***	0.169***	0.141***	0.147***	0.156***	0.137***	0.129***
	(6.62)	(5.74)	(5.62)	(4.20)	(5.17)	(4.89)	(5.32)	(5.57)	(4.60)	(3.48)
Southeast	0.142***	0.190***	0.153***	0.150***	0.146***	0.130***	0.129***	0.121***	0.0945***	0.0945**
	(5.62)	(4.62)	(4.24)	(4.63)	(4.33)	(4.78)	(5.29)	(4.51)	(3.47)	(2.55)
Southnorth	0.214***	0.280***	0.235***	0.223***	0.214***	0.183***	0.203***	0.194***	0.174***	0.174***
	(9.87)	(6.80)	(7.30)	(6.98)	(6.69)	(6.80)	(8.46)	(7.42)	(6.29)	(5.44)
Green	0.575***	0.447***	0.517***	0.597***	0.636***	0.601***	0.549***	0.497***	0.427***	0.287***
	(10.57)	(5.02)	(5.13)	(7.41)	(8.48)	(8.88)	(7.58)	(8.28)	(7.46)	(3.88)
FAR	0.0042	0.0072	0.0150***	0.0102**	0.0048	-0.0029	-0.0082*	-0.0155***	-0.0223***	-0.0221***
	(0.99)	(1.11)	(3.35)	(2.05)	(1.00)	(-0.59)	(-1.75)	(-4.03)	(-5.61)	(-3.03)
WaterView	0.215***	0.0775	0.218***	0.150***	0.176***	0.165***	0.137***	0.0983**	0.0760*	-0.0109
	(6.54)	(0.66)	(4.60)	(3.08)	(2.83)	(4.19)	(3.88)	(2.36)	(1.73)	(-0.29)
HighRise	0.0770***	0.130***	0.0983***	0.0936***	0.0947***	0.103***	0.114***	0.114***	0.0741***	-0.0191
	(6.85)	(7.67)	(5.81)	(5.85)	(6.44)	(7.31)	(8.16)	(8.66)	(4.88)	(-1.13)
NearSubway	0.0371***	0.0735***	0.0662***	0.0468***	0.0295***	0.0198**	0.0148*	0.0165**	0.0147*	0.0239**
	(4.26)	(4.86)	(5.70)	(4.33)	(2.98)	(2.30)	(1.68)	(1.97)	(1.79)	(1.97)
NearPark	0.103***	0.179***	0.163***	0.120***	0.0868***	0.0799***	0.0888***	0.0711***	0.0613***	0.0510**
	(5.57)	(4.56)	(7.92)	(7.60)	(4.14)	(4.96)	(5.37)	(4.32)	(3.51)	(2.10)
ringRoad02	1.662***	1.893***	1.839***	1.821***	1.801***	1.736***	1.688***	1.609***	1.521***	1.424***
	(73.77)	(55.65)	(62.10)	(60.98)	(58.56)	(57.07)	(56.69)	(55.71)	(41.49)	(41.09)
ringRoad23	1.541***	1.692***	1.705***	1.708***	1.674***	1.625***	1.586***	1.511***	1.399***	1.265***
	(81.86)	(48.71)	(62.20)	(61.77)	(61.22)	(56.66)	(56.46)	(63.14)	(46.05)	(44.93)
ringRoad34	1.482***	1.660***	1.622***	1.618***	1.592***	1.532***	1.503***	1.449***	1.344***	1.230***
	(94.21)	(72.74)	(72.74)	(67.36)	(62.58)	(59.37)	(56.61)	(60.23)	(45.75)	(50.63)
ringRoad45	1.324***	1.526***	1.512***	1.516***	1.474***	1.405***	1.342***	1.268***	1.145***	1.015***
	(87.53)	(60.88)	(64.33)	(67.44)	(60.47)	(56.78)	(51.75)	(53.05)	(38.91)	(42.23)
ringRoad56	0.964***	1.087***	1.081***	1.076***	1.057***	1.016***	0.996***	0.966***	0.902***	0.830***
	(67.90)	(43.96)	(52.68)	(46.68)	(42.98)	(39.17)	(37.54)	(37.65)	(29.71)	(32.42)
Constant	3.217***	2.480***	2.648***	2.765***	2.926***	3.112***	3.253***	3.436***	3.747***	4.117***
	(95.09)	(41.32)	(47.88)	(51.31)	(56.57)	(65.05)	(65.66)	(75.83)	(76.04)	(90.28)
N	8511	8511	8511	8511	8511	8511	8511	8511	8511	8511
adj./pseudo R^2	0.809	0.588	0.585	0.586	0.587	0.588	0.591	0.594	0.595	0.602

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The standard errors for quantile regression are obtained through 500 bootstrap replications.