

Housing Affordability and Emerging Economy in Urban Beijing

Jing Huang^{1, a, *}

¹Shanghai United International School, Shanghai, China

^ak827473016@outlook.com

*Corresponding author

Keywords: Housing Affordability, Econometrics Modeling, Business Analysis

Abstract. *Purpose* – The purpose of this paper is to examine determinants of housing affordability in Beijing and explore how housing affordability will influence business development, with a special focus on startups and skill-intensive firms. Empirical research is based on Beijing grid-cell level data from 2012-2014. *Methodology* – This paper firstly applies econometrics modeling techniques including OLS model, Logit model, panel regression and negative binominal model to test determinants of housing affordability and its impact on business development. Additionally, we also employ machine learning methods such as k-means clustering to classify grid cells of Beijing metro area based on some social-economic characteristics.

1. Introduction

Since the start of 21st century, cities in China began to develop in a rapid pace. Beijing, as the capital city of China, experienced enormous changes and development in various aspects. Since Beijing became the host of 2008 summer Olympics game, the city government has invested billions of capitals in the construction of stadiums, transportation system and infrastructure. People in less developed cities saw opportunities in this vast investment. Job opportunities and holistic city planning attracted massive internal migrants. There were mainly two types of internal migration: rural-urban migration and city-city migration. Rural migrants moved into cities in seek of higher wages and better working conditions, but their lack of skills limited them as workers in low value-added industries. Meanwhile, city-city migrants moved to large cities and worked as white-collar and small entrepreneurs for better career development. In 2008, the number of internal migrants peaked at nearly 5 million [1]. The influx of people leads to thriving economy, but also problems such as heated housing market and congested traffic conditions.

While massive immigration and rapid urban infrastructure development boosted the economic growth of Beijing, they also result in the deterioration of housing affordability and serious traffic congestion, which undermine the potential of sustainable development in Beijing. Beijing's housing market entered a new era in 1998 [2]. Housing affordability is commonly used to describe the difficulties related to housing issues [3]. From 1949 to 1998, China has a welfare housing scheme, in which government provides houses to people based on their lengths of service, family sizes and income levels with low rents. Hence, people did not have any financial burden for obtaining houses. However, this scheme was stopped in July 1998 to stimulate the development of market-derived housing market and economic growth. By that time, houses became goods that people spent money to purchase from real estate developers. From 2001 to 2004, housing prices were constant around 4700 RMB/m² as the industry was at its infant stage [4]. Government's reform of land policies and people's expanding demands for houses then encouraged development of the market. 2004-2007 was the period when Beijing's housing prices increased drastically by 124% [5]. Rapid economic development and increasing housing demand are major drivers of price. Also, the manipulation from real estate developers and the government's land-financing strategy accelerated the process. After that, Beijing city government announced and implemented strict regulations to cool down the housing market. However, prices consistently rose and arrived at 40592 RMB/m² in 2014.

Citizens could hardly afford the extortionate prices of buying an apartment. Immigrants also faced the policy restriction of hukou (registered permanent residence), which made them ineligible to

purchase housing and turned to rental markets. However, there was limited rental housing supply. More seriously, many of them were located in urban villages, causing immigrants to face job-housing imbalance as they worked in cities and lived in suburbs. The unideal conditions of houses also worsen the living standards of tenants.

Rental market grew significantly when online agencies occurred. In the past, house owners spread the information through agent companies in their districts. Tenants could only access the information when they saw the post of housing information. Yet, with the appearance of online agents, information became more accessible and rental turnover rate increased. Nonetheless, certain agents deliberately controlled the supply of available rental housing to raise the prices. Due to the shortage of rental housing units, group renting emerged at that time. For people working in low-paid jobs, renting a house wholly was still a huge drain on their limited income.

On the other hand, Beijing's public transit system became more mature and advanced in the past decade. In preparation for the 2008 Olympics, Beijing invested over 63 billion RMB in operating line No.5 and No.13. Moreover, commuting time was shortened as more metro lines were put into operation, and thus, increasing productivity of the labor force. Passengers then only need one card to use all public transportation tools in the city. The advancement in technology again enhanced the conveniences of using public transit. In spite of this, Beijing's metro system is not spatially evenly distributed. For instance, the current metro lines rarely connect Yanqing County with the rest areas of Beijing. Consequently, people living in Yanqing County can barely enjoy the benefits of public transit. Through 2008 to 2015, Beijing opened 7 new subway lines (No. 4,15,9,8,6,10,14), aiming at reducing congestion in ground traffic and shortening commuting time for citizens. There were policies to push the use of public transit system as well. For example, employees in state-owned enterprises could submit an expense account to the treasurer's office and get their money spent on public transit back.

Housing affordability has been a popular global topic in the 21st century as it influences talents flow, division of urban social stratum, city inclusiveness and future growth potential. Beijing now encourages formation of start-up businesses and emerging industries, but the sky-rise prices terrified young people with no financial reserves. Instead, they move to smaller cities, where resources are more accessible and affordable. In long term, Beijing may lack progressive talents and face stagnation in social and economic development.

2. Background

2.1 Social-economic Background

As the housing market was in the planned to market transition, Beijing government carried out a series of policies on the integration of land resources. The first action was to clear out the dilapidated houses in the urban areas that were built in the 1960s, so more land could be sold to real estate developers through government auction for the construction of new housing. This leads to forcible movement of residents in those houses. Most of them were low-end population, who could hardly afford purchasing an apartment in the city. Thus, they moved to suburb areas. The population structure of the inner city then shifted to be more educated and financially capable economically actives. The integration of land resources helped Beijing to form a more orderly city, but it would reduce affordable rental housing supply (result in the soar of housing price) in the future because the control was in real estate developers' hands.

To tackle this issue, Beijing launched public rental housing program in 2012 [6]. At first, it was known as the public rental housing and economic housing program to provide affordable rental housing to citizens. However, the limitations of this program outweighed its benefits: the supply was limited; location of the houses were in periurban areas, which was unideal for employees working in inner city; most importantly, only local citizens or people working in Beijing for many years were qualified to apply for public rental housing program. Therefore, this program did not provide effective solutions to satisfy the large rental housing demand. Young immigrants with no access to those public rental housing and also cannot afford the expansive housing in urban area had to live in

suburb areas and suffer from the long commute time, or turn to group rent with other people. Later, both Beijing city government and private developers developed some affordable rental housing projects such as youth apartments, targeting at young talents without Beijing hukou (residence cards). This encouraged talented people to stay and work in Beijing. Lastly, long leasing apartments were introduced to provide housing for people who expect to work in Beijing for a long period of time with a relatively affordable price.

One major consequence of lack of affordable rental housing units is the Job-housing imbalance. In the urban area of Beijing, city planners chose certain districts and made economic development as their priorities, such as Zhaoyang and Haidian. In 2015, these two districts achieved a total GDP of 925 billion RMB, composing 40% of the city's overall GDP [7]. Therefore, the land use in these areas were prudently planned. In order to maintain the thriving economy, land was utilized for large companies, especially multinational companies, to build their offices. There were also office buildings for renting, targeting at start-up businesses or small enterprises. Therefore, no space was left for the construction of housing. In addition, lands in these areas were extremely expensive, even for real estate developers in the acquisition process. Subsequently, they moved away from urban area and constructed houses there. Then all the available housing was far away from office areas. Young people with limited financial capacity have to live in suburb areas where rental price is lower. Consequently, job-housing imbalance emerged.

This resulted in intolerably and frustrating long commuting, which seriously undermines working efficiency. Among all metropolis in China, Beijing had the longest commuting distance: 19.2 kilometers. More interestingly, the percentage of people who had a commuting distance within 5 kilometers was 31.8%, whereas the percentage of people who had a commuting distance over 25 kilometers was 17.1% and the average commuting time of 56 minutes (highest in China). This clearly confirms the severity of job-housing imbalance [8]. Yao and Wang's work shows that people tend to commute by public transportation when distance becomes longer [9].

Currently, Beijing government encouraged the mass innovation and entrepreneurship. However, unaffordable commercial and residential space became the largest challenge for young entrepreneurs. Some start-ups had to leave for more affordable cities, which caused Beijing to lose great talents and future potential for innovation-driven economic development.

2.2 Research Background

It is very crucial for Beijing's future sustainable development to have a better knowledge of spatial distribution of affordable housing, public amenities, and their impacts on talents and innovation. More specifically, we explore determinants of housing affordability in Beijing urban area. This paper is featured with two contributions to the literature.

First, we test housing affordability determinants at grid-cell level, which is a very detailed geographical level in China. Few previous papers conduct empirical research at grid-cell level. One major reason is the lack of data. Additionally, we focus on rental housing market instead of the whole housing market. Since rental housing market is more related to welfare of immigrants, our empirical results can more effectively provide policy implications on how to make Beijing an inclusive city. After providing empirical evidence on determinants of housing affordability at grid-cell level, this paper further studies how it will influence business development and the prosperity of innovation-driven economy. This part makes a meaningful attempt to explore the social economic consequences of soared housing price and how entrepreneurs incorporate housing affordability in their business location choice, and the trade-off between it and agglomeration of business resources.

Next, our paper uses traditional econometrics including OLS model, Logit model, panel regression and negative binomial model and modern machine learning method such as clustering and random forest tree to articulate factors influencing housing affordability.

Then we apply machine learning method to classify grid cells within Beijing metro area into several categories in terms of housing affordability, public amenities richness and diversity, business resources, public transit accessibility. Subsequently, we compare the developments of Beijing's

metro lines and innovation-driven economy to explore how geographical variation in housing affordability is related to spatial imbalance of innovation-driven firm agglomeration.

3. Data Structure

The data set used in this paper is grid-cell panel data, with 2547 grid cells in total ranging from 2012 to 2014. The data divides Beijing metro area into 2547 grid cells. Each grid cell is sized about 1 km * 1 km.

Within each grid, a range of variables is collected to investigate their impacts on housing affordability and development of innovation-driven entrepreneurs. The distribution of some variables is very dispersed as each grid cell varies significantly in its level.

We use four main indicators to measure within-city housing affordability from two dimensions: area and number of units, either of them recorded using two methods. The first one is the number of affordable housing units in a grid cell. This data gathered from 5i5j indicates the available number of houses that are in the market and open to tenants. The second one is area of affordable housing. The average rental housing area in Beijing is 69.90m², according to 5j5i.com. A rental housing with area below 69.90m² is typically considered as affordable. The data uses two methods to measure housing affordability. One is to use price-income ratio, which divides the income of a household by the price of a rental housing. A ratio below 30% is judged as affordable. The other is to use the average deflated rental price, which sets 2005 as the base year. A rental price below 28.87 per m² is deemed as affordable. It is worth noticing that measurement of housing affordability is the dependent variable for the housing affordability model. Meanwhile, it is the primary independent variable in business development model.

There are altogether 7641 observations of the number of affordable housing units (num_afford_m1) and average rental housing area (area_afford_m1). The number of affordable housing units has a mean of 13 units, a median of 0 unit and a range of 388. These summary statistics show an extremely dispersed distribution, for over 50% of the grid cells do not have any affordable housing units whereas one specific grid cell can have 388 affordable housing units. The average rental housing area has a mean of 621 m², a median of 0 m² and a range of 19731 m². This variable also has a significantly dispersed distribution. By analyzing the two variables, it's certain that the majority of affordable housing is concentrated in a small number of grid cells. We will employ some remedy measures to mitigate its dispersed distribution in later modeling stages.

Below are several important independent variables that can we include in empirical models:

1) Presence of state-owned enterprises

This can be demonstrated by characteristics such as state-owned enterprise output (SOE_output) and share of state-owned enterprises' employees of the permanent residents (SOE_emp_sh) in the grid cell. The former characteristic exhibits the share of output from state-owned enterprises and the latter one displays the percentage of permanent residents within the grid cell employed in state-owned enterprises. The data is gathered from the State Administration of Industry and Commerce. It is momentous to explore the relationship between presence of state-owned enterprises and housing price because state-owned enterprises positively propel an area's economy, which in turn affects the value and zoning of the land. We expect that the presence of state-owned enterprises may have positive impact on the grid-cell level housing affordability because many SOEs have implemented some policies to improve their employees' welfare.

2) Distance to Beijing CBD area

It measures the geographical distance from the centroid of this grid cell to Beijing CBD area. It is meaningful to explore the relationship between distance to urban center and housing price. We expect that housing affordability may diminish as the distance to CBS increases. We also put a polynomial function of distance to CBD (add second-order term) to explore whether the distance to CBD has a non-linear effect on housing affordability.

3) *Richness and diversity of restaurants and entertainment amenities*

Richness of restaurants and entertainment amenities measures the total number of restaurants and entertainment amenities within a grid cell. Diversity of restaurants and entertainment amenities measures the amount of different cuisine types and styles of restaurants and entertainment amenities a grid cell contains. The measurements are made based on information obtained from dianping.com (the Chinese version of Yelp.com). The data records roughly 78,454 restaurants in total in the period of 2013-2014. It is evocative to explore the relationship as areas with more diverse restaurants and entertainment amenities can better satisfy people's wants and improve their living standards. These grid cells may then have more affordable housing as real estate developers know people have a preference of residing here.

4) *Nightlife Density*

It measures the level of activity occurred in a grid cell in the evening, when people get off work and enjoy their social lives. Satellite imagery from the Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime sensor through National Aeronautics and Space Administrations (NASA) Global Imagery Browse Services (GIBS) is utilized to unveil the nightlife density. Mean nighttime light intensity values for each grid cell is calculated from the satellite image with the BMA shapefile. These values are used as a proxy for economic activity at grid-cell level. It is important to explore the relationship because areas with high level of economic activity tend to have higher land prices, thus, causing high rental housing prices.

5) *Business resources: firm stock*

It measures the total number of firms registered within the grid cell. This data is acquired from the State Administration of Industry and Commerce. This variable measures the business development of this grid-cell. With more firms, the grid cell is likely to have higher GDP. This may lead to higher land prices and higher rental housing prices. However, grid-cells with lots of firms may be a more affordable area because firms may be attracted to place with lower rental price. Additionally, housing units near business district may be old apartments which are less popular than newly developed communities. Also, we construct two more specific measures for firm stock: number of skill-intensive firms and number of innovation-driven firms based on average education attainment and the extent of applications of new technology at the industry level.

6) *Number of universities and research institutions*

All the universities and research institutions in Beijing is collected to approximate human capital pools within the city. The Beijing Municipal Commission of Education and the Ministry of Education of the People's Republic of China provide the list of universities and research institutions. The branches and divisions of the Chinese Academy of Sciences in Beijing are counted as the research institutions. This is because the R&D activity in these institutions are relevant for entrepreneurial activity. Larger number of universities and research institutions results in higher concentration of talents. Consequently, there is higher demand of rental housing. Therefore, the housing prices will be higher.

As an independent variable in the business development model, housing affordability will influence innovation-driven economic development. Firm-stock measures the total number of firms existing in a grid-cell. The number of start-up businesses measures the number of firms that operated in Beijing for less than five years in a grid cell.

Grid-cell data provides an extremely detailed scope for analyzing housing affordability. However, grid-cells are interlinked and they influence each other. Hence, discounted measure of independent variables is generated to reflect such influences. This is achieved by using the travel time matrix calculated from Baidu API. Since we have values for each grid-cell at each variable, we use the travel time (minimum between subway travel time and walk time) from one grid cell to each of the all other grid cells to discount the value of these variables. For example, if one grid cell is recorded to have 10 restaurants and its takes 10 minutes to travel from my grid-cell to that cell, then my grid cell can get

1(10/10) restaurant from this cell. This measure is basically a accessibility measure. In the final step, we aggregate all those discounted measures from my grid cell to all others to generate the “accessibility” version of variables.

Nevertheless, there are certain limitations in the data structure. Though grid-cell level data can showcase variables at an extensively detailed extent, social-economic and built environment data such as average education level, income, household structure, age, area of green space and availability of public services is not available at such detailed geographical level. Census data collection is mainly conducted at the city and province level and we lack data at grid-cell level.

Table 1. Summary statistics

Variables	Number of observations	Mean value	Standard deviation	Minimum value	Maximum value
Number of affordable housing units within the grid cell by metod 1	2547	12.79	35.29	0	388
Number of affordable housing units within the grid cell by metod 3	2547	20.67	60.76	0	669
Area of affordable housing units within the grid cell by metod 1	2547	620.59	1739.65	0	19730.6
Area of affordable housing units within the grid cell by metod 3	2547	1131.82	3347.07	0	37143.62
Distance to CBD	2547	22.4	10.33	0.25	49.87
Employment share from SOE sector	2547	0.38	0.32	0	0.98
Assest share from SOE sector	2547	0.38	0.34	0	0.88
Outputshare from SOE sector	2547	0.35	0.33	0	0.98
Number of startups within the grid cell	2547	14.29	36.35	0	576
Number of innovation spaces within the grid cell	2547	0.29	1.44	0	30
Number of subways lines within the grid cell	2547	0.14	0.43	0	4
Number of research institutions within the grid cell	2547	0.05	0.48	0	14
Number of universities within the grid cell	2547	0.03	0.21	0	3
Number of restaurants within the grid cell	2547	30.8	60.33	0	647
Diversity of restaurants within the grid cell	2547	1.03	0.89	0	2.53
Average nightlife density within the grid cell	2547	19.69	15.45	0	227.17

4. Empirical Method

For the empirical methodology, we firstly apply econometrics modeling to estimate grid-cell level housing affordability using some selected independent variables. Econometrics models included in this paper: OLS, Logit, fixed-effect panel regression and negative binominal regression. But econometrics methods heavily rely on building a structured model instead of data mining. So then we also apply K-means clustering to classify grid cells from these dimensions: livability, level of business development and prosperity of entrepreneurship and innovation, and explore relationship among these clustering results.

For econometrics modeling, we treated our data as pooled data because the time window is relatively short – from 2012 to 2014. This treatment is thus reasonable. Dependent variable used in econometrics modeling is housing affordability. Independent variables included in econometrics modeling are: presence of state-owned enterprises, distance to Beijing CBD area, richness and diversity of restaurants and entertainment amenities, nightlife density, business resources (firm stock) and number of universities and research institutions.

First, we used OLS for preliminary analysis. OLS explores the linear relationship between housing affordability and each independent variable so that we could have a rough exploration of structural relationship.

Next, we used Logit model which is employed when the dependent variable is binary (0/1). Hence, we transformed the continuous variables (number/area of affordable housing units) into a dummy variable based on a cut-off value and then utilized logit (0-1/dummization) to account for the dispersion and some outliers in the data. Dummy transformation results in loss of information and a much smaller variance of the dependent variable.

To further mitigate the dispersion of the dependent variable and retain all variation, we used the negative binomial model, which is very effective at count variables, large variance and dispersed distribution. The model customized for the count variables (or dispersed variable) measured housing affordability.

Since we have panel data, panel regression is applied to add fixed effects and control for time variation. We added two-way fixed effects: grid-cell fixed effect and year fixed effect.

On the other hand, we used machine learning algorithm: k-means clustering. Instead of classifying grid-cells from geographical perspective, we wanted to classify grid cells within Beijing metro area into several clusters based on data mining of important indicators. We divided grid cells into different clusters based on their livability, level of business development and prosperity of entrepreneurship and innovation. Then we explored how these clustering results interlink with each other.

5. Empirical Analysis and Conclusion Econometrics

5.1 Modeling

For all models, we first looked at its p-value to determine the significance of each independent variable. Here we use 0.1 significance level for p-value criteria. More specifically, determining whether the variable has significant effects (the regression coefficient is significantly different from 0) on the outcome variable depends on its whether the p-value of its coefficient is below 0.1. For independent variables with significant effect, we evaluate the degree of influence based on the effect size of its coefficient. The larger the significant coefficient, the more influential this variable on the outcome. Additionally, the positive sign of regression coefficient indicates a positive effect and the negative sign implies the negative effect.

num_afford_m1	Robust					[95% Conf. Interval]
	Coef.	Std. Err.	t	P> t		
d_cbd_2	.018079	.0052625	3.44	0.001	.0077597	.0283983
d_CBD	-1.314543	.3019513	-4.35	0.000	-1.906647	-.7224379
SOE_emp_sh	38.08777	10.40246	3.66	0.000	17.68931	58.48624
SOE_asset_	-25.45684	12.60953	-2.02	0.044	-50.1832	-.7304699
SOE_output	-16.95514	7.185574	-2.36	0.018	-31.04553	-2.864753
c_subway	.5605842	2.220536	0.25	0.801	-3.793726	4.914895
c_ri	-1.089465	1.754186	-0.62	0.535	-4.529297	2.350367
c_university	-2.255355	4.058164	-0.56	0.578	-10.21312	5.70241
n10	.215972	.0412948	5.23	0.000	.1349959	.296948
d10	4.892468	.9795064	4.99	0.000	2.971727	6.813209
nightlife_average	.0280184	.0397396	0.71	0.481	-.0499082	.105945
Firmstock	.0039485	.0029751	1.33	0.185	-.0018854	.0097824
_cons	18.69196	4.074808	4.59	0.000	10.70156	26.68236

Fig. 1. Housing affordability model 1

We start by using OLS model to explore determinants of grid-cell level housing affordability. In the OLS model for number of affordable housing units within the grid cell by method 1 (priceincome ratio), employment share from SOE sector has a great significance as its p-value is 0.000. It has a coefficient of 38.09, which showcases as employment share from SOE sector increases by 1, the number of affordable housing will increase by 38.09. This is because in areas with large amount employees employed in SOE, SOEs will try to increase employee welfare by providing them affordable housing. This statistical result corresponds to the theoretical implications.

Number of research institutions and universities within the grid cell seem to have trivial effects on housing affordability, which is slightly out of expectation because we think that area near educational facilities may have more affordable housing units provided for young students. Grid cells with richer and more diverse restaurants also seem to be better equipped with affordable housing units. However, the density of night-life activities is weakly related to housing affordability.

Meanwhile, the number of subway lines within the grid cell is the least significant factor among all in this model. This is potentially due to the fact that areas with more subway lines offer people greater convenience in commuting, thus, have a higher land price. This causes unaffordable housing.

num_afford_m3	Robust					[95% Conf. Interval]
	Coef.	Std. Err.	t	P> t		
d_cbd_2	.0364177	.0095697	3.81	0.000	.0176522	.0551833
d_CBD	-2.45448	.5568938	-4.41	0.000	-3.546509	-1.362452
SOE_emp_sh	18.84853	11.84644	1.59	0.112	-4.381493	42.07855
SOE_asset_	-6.844644	21.85624	-0.31	0.754	-49.70314	36.01385
SOE_output	-18.4573	14.44779	-1.28	0.202	-46.78836	9.873759
c_subway	3.810256	4.574594	0.83	0.405	-5.160191	12.7807
c_ri	-4.859493	3.43266	-1.42	0.157	-11.59069	1.871704
c_university	-4.120772	7.230957	-0.57	0.569	-18.30015	10.05861
n10	.3329111	.0709833	4.69	0.000	.193718	.4721042
d10	6.970982	1.63009	4.28	0.000	3.774493	10.16747
nightlife_average	.1297289	.077784	1.67	0.095	-.0228	.2822577
Firmstock	.0067961	.0043509	1.56	0.118	-.0017356	.0153278
_cons	33.31258	7.972471	4.18	0.000	17.67914	48.94602

Fig. 2. Housing affordability model 2

Number of affordable housing units within the grid cell by method 3 (average deflated rental price) is also run in OLS model, but it offers different significant factors as compared to number of affordable housing units within the grid cell by method 1. In the previous model, nightlife density is insignificant, but it is significant in this model, with a positive coefficient of 0.130. The reason is that areas with higher nightlife density tend to have more economic and human activities, more residents and more livable and richer amenities. Good built-environment factors, such as green space, also aid to promote the rental housing price.

Then we transform the dependent variable into a binary variable by setting a cutoff value and apply logit model to compare results. In the Logit model of affordable housing by method 1 (income-rent ratio), number of restaurants within the grid cell and diversity of restaurants within the grid cell are significant. They both have a positive coefficient: 0.013 and 2.282, respectively. This results can be explained by the fact that when more types of restaurants are built, people have more choices for dining. This makes certain grid cells more attractive to residents. Hence, real estate developers constructed more affordable housing in these areas. Noticeably, we capture a non-linear effect of the distance to CBD on the housing affordability. Housing affordability will increase as the distance to CBD increases up to 10-15 kms and then decrease. The share of state-owned enterprises employees to in local permanent residents still has positive effects on housing affordability in logit model. Firm stock has a marginally significant positive effect on housing affordability. Grid cells with relatively rich business resources tend to have slightly more affordable housing units than other cells. Number of research institutions and universities remain insignificant in the logit model.

Affordable_m1	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
d_cbd_2	-.0069227	.0013935	-4.97	0.000	-.0096539	-.0041915
d_CBD	.1396069	.0486878	2.87	0.004	.0441806	.2350331
SOE_emp_sh	2.178085	.658629	3.31	0.001	.8871962	3.468974
SOE_asset_	-.8055772	1.336718	-0.60	0.547	-3.425496	1.814342
SOE_output	-2.387302	1.109374	-2.15	0.031	-4.561635	-.2129696
c_subway	.0902953	.1439903	0.63	0.531	-.1919204	.372511
c_ri	-.0693961	.1961135	-0.35	0.723	-.4537715	.3149794
c_university	-.3044078	.2781833	-1.09	0.274	-.8496371	.2408215
n10	.0131038	.0038037	3.44	0.001	.0056486	.0205589
d10	2.281861	.3123877	7.30	0.000	1.669592	2.894129
nightlife_average	.0003198	.0079157	0.04	0.968	-.0151947	.0158342
Firmstock	.0003265	.0001717	1.90	0.057	-1.00e-05	.000663
_cons	-5.990384	.6629673	-9.04	0.000	-7.289776	-4.690992

Fig. 3. Housing affordability model 3

Affordable_m3	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
d_CBD	-.0867813	.0113893	-7.62	0.000	-.1091039	-.0644586
SOE_emp_sh	2.933725	.6488175	4.52	0.000	1.662066	4.205384
SOE_asset_	-2.766326	1.341763	-2.06	0.039	-5.396133	-.1365189
SOE_output	-.4502293	1.105319	-0.41	0.684	-2.616615	1.716157
c_subway	.0223357	.1465591	0.15	0.879	-.2649148	.3095862
c_ri	-.0266851	.1292375	-0.21	0.836	-.2799859	.2266157
c_university	.0821856	.2612031	0.31	0.753	-.4297631	.5941344
n10	.0064135	.0033992	1.89	0.059	-.0002489	.0130759
d10	2.787087	.288702	9.65	0.000	2.221242	3.352933
nightlife_average	.0170955	.0063908	2.68	0.007	.0045698	.0296212
Firmstock	.0005528	.0002323	2.38	0.017	.0000975	.0010082
_cons	-5.610899	.5726901	-9.80	0.000	-6.73335	-4.488447

Fig. 4. Housing affordability model 4

Then we still use the logit model but using the measurement of housing affordability defined by average price to test the effects from determinants. Number of firms with the business entity registration within the grid cell is significant. Yet, the size of its coefficient is only 0.00055, a very small number. The positive relationship is because as areas enter the grid cell, they bring in more investment. This amount of money would help the grid cell to be more advanced and descent for living. Therefore, more affordable housing is available.

Next, we used fixed effect panel regression to explore determinants of number of affordable housing units within the grid cell by method 1 (price-income ratio). In this model, distance to CBD, employment share from SOE sector, output share from SOE sector, number of restaurants and diversity of restaurant maintain significant. However, asset share from SOE sector becomes insignificant as compared to OLS model. This is probably because when grid cell is fixed, there are fewer grid cells with high asset share from SOE sector have affordable housing. Thus, the significance of its impact decreases.

num_afford_m1_d	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
d_CBD	-.3834945	.0706447	-5.43	0.000	-.5219557	-.2450334
SOE_emp_sh	30.54286	9.76745	3.13	0.002	11.39901	49.68671
SOE_asset_	-15.63395	12.0916	-1.29	0.196	-39.33305	8.06515
SOE_output	-19.78285	7.036516	-2.81	0.005	-33.57417	-5.991537
c_subway	.7067738	2.183396	0.32	0.746	-3.572603	4.98615
c_rriuniver	-1.47195	1.621113	-0.91	0.364	-4.649273	1.705373
n10_d	.2609504	.0132251	19.73	0.000	.2350296	.2868711
d10_d	5.853516	.5011564	11.68	0.000	4.871268	6.835765
nightlife_average	.0086977	.0486905	0.18	0.858	-.086734	.1041294
Firmstock_d	1.70e-07	4.75e-06	0.04	0.971	-9.13e-06	9.47e-06
_cons	7.743946	2.476456	3.13	0.002	2.890181	12.59771
sigma_u	29.239499					
sigma_e	.02616171					
rho	.9999992	(fraction of variance due to u_i)				

Fig. 5. Housing affordability model 5

num_afford_m3_d	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
d_CBD	-.4716743	.1324232	-3.56	0.000	-.7312191	-.2121296
SOE_emp_sh	2.865105	11.4427	0.25	0.802	-19.56217	25.29238
SOE_asset_	17.3732	21.25737	0.82	0.414	-24.29048	59.03688
SOE_output	-26.35677	14.47276	-1.82	0.069	-54.72286	2.009312
c_subway	1.27171	4.721903	0.27	0.788	-7.983051	10.52647
c_rriuniver	-6.810273	3.640025	-1.87	0.061	-13.94459	.3240441
n10_d	.5578882	.0238157	23.43	0.000	.5112103	.6045661
d10_d	5.367002	.8889708	6.04	0.000	3.624651	7.109353
nightlife_average	-.0395093	.0905216	-0.44	0.663	-.2169283	.1379097
Firmstock_d	-.0000423	.0000108	-3.93	0.000	-.0000634	-.0000212
_cons	10.76407	4.714829	2.28	0.022	1.523174	20.00496
sigma_u	51.801445					
sigma_e	.04498146					
rho	.99999925	(fraction of variance due to u_i)				

Fig. 6. Housing affordability model 6

Panel regression is also applied to the number of affordable housing units within the grid cell by method 3 (average deflated rental price). Distance to CBD, employment share from SOE sector, output share from SOE sector, number of restaurants and diversity of restaurant still show strong significant. However, nightlife density becomes insignificant as compared to OLS model. This is possibly due to the fact when nightlife density is high, economic and human activities increase, which then result in higher land prices. This made the areas unaffordable.

Finally, we employed negative binomial model. We first investigate the level of significance of different independent variable of the number of affordable housing units within the grid cell by method 1 (price-income ratio) and the number of affordable housing units within the grid cell by method 3 (average deflated rental price).

num_afford_m1	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
d_cbd_2	-.0047319	.0010626	-4.45	0.000	-.0068146	-.0026492
d_CBD	.1078464	.0386277	2.79	0.005	.0321376	.1835552
SOE_emp_sh	2.038942	.6054694	3.37	0.001	.8522437	3.22564
SOE_asset_	-1.506103	1.05959	-1.42	0.155	-3.582862	.5706555
SOE_output	-1.583137	.9255473	-1.71	0.087	-3.397176	.2309025
c_subway	.123697	.1017868	1.22	0.224	-.0758013	.3231954
c_ri	-.0292598	.0468283	-0.62	0.532	-.1210415	.0625219
c_university	-.2569572	.1458245	-1.76	0.078	-.5427679	.0288535
n10	.0076518	.0012639	6.05	0.000	.0051747	.010129
d10	1.972607	.1463141	13.48	0.000	1.685836	2.259377
nightlife_average	.0283266	.0091864	3.08	0.002	.0103216	.0463316
Firmstock	.0000939	.0000643	1.46	0.144	-.0000322	.0002201
_cons	-2.219166	.4645684	-4.78	0.000	-3.129703	-1.308628
/lnalpha	1.391985	.0675963			1.259498	1.524471
alpha	4.022826	.271928			3.523654	4.592713

Fig. 7. Housing affordability model 7

num_afford_m3	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
d_cbd_2	-.0037235	.0010325	-3.61	0.000	-.0057472	-.0016997
d_CBD	.0691789	.038551	1.79	0.073	-.0063797	.1447375
SOE_emp_sh	1.782625	.6979632	2.55	0.011	.4146422	3.150608
SOE_asset_	-.9004211	.9736615	-0.92	0.355	-2.808763	1.00792
SOE_output	-1.580804	.8573991	-1.84	0.065	-3.261275	.0996671
c_subway	.1900071	.1343474	1.41	0.157	-.073309	.4533232
c_ri	-.0598469	.0500151	-1.20	0.231	-.1578747	.0381809
c_university	-.2891409	.1467913	-1.97	0.049	-.5768466	-.0014352
n10	.0074851	.0015357	4.87	0.000	.0044751	.010495
d10	2.369471	.1479298	16.02	0.000	2.079534	2.659408
nightlife_average	.0360001	.0090871	3.96	0.000	.0181896	.0538106
Firmstock	.0001238	.0000764	1.62	0.105	-.0000259	.0002735
_cons	-2.673157	.4397437	-6.08	0.000	-3.535039	-1.811276
/lnalpha	1.424905	.0619189			1.303546	1.546263
alpha	4.157461	.2574255			3.68233	4.693898

Fig. 8. Housing affordability model 8

The majority of independent variables that showcase high level of significance in previous models are still highly significant, including distance to CBD, number of restaurants, diversity of restaurants and nightlife density.

Moreover, we used negative binomial model to explore the determinants of number of firms with the business entity registration within the grid cell. The number of affordable housing units within the grid cell by method 3 (average deflated rental price) is a significant factor. This is because we want to examine the effect of housing affordability on the number of firms with the business entity registration within the grid cell. It is shown that the number of affordable housing units measured by method 3 (average deflated rental price) is strongly significant. This can be explained by the theory that areas with more affordable housing have higher level of economic and human activities. In addition, areas with more affordable housing attract more residents, which then became a potential source of labour force for the firms. Therefore, firms would like to locate in those areas. Moreover, the number of restaurants, diversity of restaurants and nightlife density are significant factors as well. They are all factors related to the level of economic and human activities and welfare of employees.

Firmstock	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
num_afford_m3	.0015133	.0005107	2.96	0.003	.0005124	.0025141
d_cbd_2	-.0007314	.000495	-1.48	0.140	-.0017016	.0002388
d_CBD	.0297962	.0227592	1.31	0.190	-.0148111	.0744034
SOE_emp_sh	-.2955216	.457515	-0.65	0.518	-1.192235	.6011914
SOE_asset_	-.1110681	.8195794	-0.14	0.892	-1.717414	1.495278
SOE_output	.5368289	.7826173	0.69	0.493	-.9970727	2.070731
c_subway	.0402393	.0865789	0.46	0.642	-.1294521	.2099308
c_ri	-.0441618	.0388733	-1.14	0.256	-.120352	.0320284
c_university	-.1324627	.118498	-1.12	0.264	-.3647146	.0997891
n10	.0059323	.0011237	5.28	0.000	.0037299	.0081347
d10	1.332683	.0955685	13.94	0.000	1.145373	1.519994
nightlife_average	.024967	.0075161	3.32	0.001	.0102358	.0396983
_cons	1.846757	.3174508	5.82	0.000	1.224565	2.468949
/lnalpha	.9481302	.0476591			.8547201	1.04154
alpha	2.580879	.1230023			2.350716	2.833578

Fig. 9. Housing affordability model 9

c_startup	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
num_afford_m1	.0040578	.001108	3.66	0.000	.0018862	.0062294
d_cbd_2	-.0008635	.0003378	-2.56	0.011	-.0015255	-.0002015
d_CBD	-.0049854	.0164508	-0.30	0.762	-.0372283	.0272575
SOE_emp_sh	-.1649532	.3308322	-0.50	0.618	-.8133724	.483466
SOE_asset_	-2.740882	.7624377	-3.59	0.000	-4.235232	-1.246531
SOE_output	3.077958	.6868988	4.48	0.000	1.731661	4.424254
c_subway	.1206321	.0526169	2.29	0.022	.0175048	.2237593
c_ri	.0287841	.0348623	0.83	0.409	-.0395448	.097113
c_university	.381304	.1206706	3.16	0.002	.1447939	.6178141
n10	.0012143	.0007138	1.70	0.089	-.0001847	.0026133
d10	.8583145	.082277	10.43	0.000	.6970545	1.019574
nightlife_average	.0299648	.0062201	4.82	0.000	.0177736	.042156
Firmstock	.0003382	.0000563	6.01	0.000	.0002279	.0004486
_cons	.3771762	.2236404	1.69	0.092	-.061151	.8155034
/lnalpha	.2773507	.0672603			.145523	.4091785
alpha	1.319629	.0887586			1.156644	1.50558

Fig. 10. Housing affordability model 10

Finally, we used negative binomial model to explore the determinants of number of startups within the grid cell. The significant factors are similar to the ones for firm stock. Housing affordability leads to higher level of economic activities, which provides startups more opportunities to survive and develop.

5.2 Machine Learning

Additionally, we use K-means clustering to classify 2547 grid cells in Beijing into four clusters from three dimensions: livability, development of State-owned enterprises / overall industrial base (i.e “traditional economy”), and prosperity of emerging economy. We conduct “livability” clustering based on five variables: number of restaurant amenities within the grid cell, diversity of restaurant amenities, number of affordable housing units measured by price-income ratio and relative price level, and number of subway stations within the grid cell. We measure development of traditional economy using total industrial assets, SOE employer share in local permanent residents, SOE total output, number of research institutions and universities within the grid cell, and the number of registered firms. We measure the condition of emerging (innovationdriven) economy using number of start-ups, number of innovation spaces, number of skillintensive firms, the number of innovation-driven firms within the grid cell. The detailed Kmeans clustering results and the characteristics of each cluster are listed below:

livability_cluster <int>	mean_n10 <dbl>	mean_d10 <dbl>	mean_number_afford_m1 <dbl>	mean_number_afford_m3 <dbl>	mean_area_afford_m1 <dbl>	mean_area_afford_m3 <dbl>	mean_csubway <dbl>
1	31.12833	1.0331959	12.96464	21.04603	626.5295	1151.2354	0.13976528
2	23.01333	0.8880322	7.20000	12.49333	342.0341	688.4572	0.09333333
3	30.02541	1.0012680	13.22123	20.76233	651.3844	1145.3797	0.11659193
4	27.72619	1.0583788	9.93750	14.86607	505.6715	824.7396	0.14285714

Fig. 11. Number of grid cells within each “Livability” cluster, Cluster 1: 2187, Cluster 2: 25, Cluster 3: 223, Cluster 4: 112

business_cluster <int>	mean_total_asset <dbl>	mean_soe_emp_sh <dbl>	mean_soe_asset <dbl>	mean_soe_output <dbl>	mean_Firmstock <dbl>
1	276432.0	0.3590784	0.3090218	0.2829241	258.0865
2	237368.2	0.4317405	0.4560390	0.4224459	279.2923
3	277653.7	0.3733673	0.3558946	0.3369908	234.2699
4	265918.2	0.3975376	0.3706996	0.3473396	252.7213

Fig. 12. Number of grid cells within each “Traditional Economy” cluster, Cluster 1: 104, Cluster 2: 284, Cluster 3: 1561, Cluster 4: 598

emerging_cluster <int>	mean_startup <dbl>	mean_space <dbl>	mean_skillfirms <dbl>	mean_innofirms <dbl>
1	14.22930	0.2860301	113.7436	71.81270
2	14.64550	0.3531746	106.7222	66.04630
3	23.58333	0.7500000	124.7500	85.16667
4	14.25287	0.3103448	147.6207	91.82759

Fig. 13. Number of grid cells within each “Emerging Economy” cluster, Cluster 1: 2262, Cluster 2: 252, Cluster 3: 4, Cluster 4: 29

From above clustering results, we find that most grid cells within Beijing metro area are featured with relatively good livability, as indicated by richness of amenities and stock of affordable rental housing units. The clustering of livability is more dispersed. In contrast, the clustering of development in traditional economy is more balanced and the difference in each measurement across clusters are relatively smaller. The clustering of prosperity of emerging economy exhibit similar spatial pattern with livability. Most grid cells has very limited presence of emerging or innovation-driven economy. But 29 grid cells achieves a good development of emerging economy.

We then recode all cluster variables based on the level of indicators in a 4 likert-point scale (the higher value indicates a higher degree of livability, development of traditional economy, prosperity of emerging economy separately). Then we calculate the spearman correlation among these variables. Here, the reason for using spearman correlation coefficient is that three clustering variables are ordinal variables. All correlation coefficients are highly significant. Livability is weakly correlated with the traditional economy, with a correlation coefficient at 0.03. Also, development of traditional economy and emerging economy is positively correlated but the effect size of correlation coefficient is only 0.11. Noticeably, livability is highly correlated with the development of emerging economy. Such difference tells us that the emerging economy values livability more than traditional economy, which is a very meaningful research finding worth more empirical research.

6. Conclusion

This paper applies econometrics modeling and machine learning analysis to explore determinants of housing affordability, the development of start-ups and structural relationship among livability, development of traditional economy and prosperity of emerging economy based on grid-cell level data of Beijing. Generally, we find that distance to CBD has a non-linear effect on the housing affordability. SOE employee share in local permanent residents, richness and diversity of amenities and nightlife density significantly influence affordability. But the economic size of SOE, presence of universities/research institutions and subway stations show a weaker effect. K-means clustering results indicate that the development of emerging economy is highly positively related with the livability and has negligible relationship with the condition of traditional economy at the grid cell level. More research is required to provide more empirical evidence on this topic.

References

- [1] http://tjj.beijing.gov.cn/nj/main/2009_ch1/content/mV33_0308.htm

- [2] Chen, Junhua, Fei Guo, and Ying Wu. "One decade of urban housing reform in China: Urban housing price dynamics and the role of migration and urbanization, 1995–2005." *Habitat International* 35.1 (2011): 1-8.
- [3] Mengjie, Liu, Richard Reed, and Hao Wu. "Challenges facing housing affordability in Beijing in the twenty-first century." *International journal of housing markets and analysis* 1.3 (2008): 275-287.
- [4] <http://data.stats.gov.cn/easyquery.htm?cn=E0105&zb=A03®=110000&sj=2018>
- [5] <http://data.stats.gov.cn/easyquery.htm?cn=E0105&zb=A03®=110000&sj=2018>
- [6] <http://fgcx.bjcourt.gov.cn:4601/law?fn=chl387s971.txt&dbt=chl>
- [7] <http://tjj.beijing.gov.cn/nj/qxnj/2016/zk/indexch.htm>
- [8] <http://travel.people.com.cn/n/2015/0128/c41570-26463574.html>
- [9] Yao, Yongling, and Shuai Wang. "Commuting tools and residential location of suburbanization: evidence from Beijing." *Urban, Planning and Transport Research* 2.1 (2014): 274-288.