Modelling Local Amenities with Online Open-source Data in a New Spatial Equilibrium Model: Insights from Applications for Beijing

Li Wan, Xiao Rong and Ying Jin

Abstract

This paper presents a new, general purpose method for modelling local amenities in a city-level spatial equilibrium model with emerging data sources. A log-form utility function is introduced to differentiate local amenities from other hard-to-observe influences of locational choice for residential and job location. In particular, we use the online open data of schools and hospitals in Beijing to improve model parameterization and calibration at high spatial resolution. The new local amenities element can improve the model's fidelity on residence location choice by over 30%, which is a step forward in decomposing the zonal attractiveness in spatial equilibrium models. Moreover the local amenities component provides a new interface for the spatial equilibrium models, where quantification of the combined effects of urban land-use and local amenities policies can be simulated on a more consistent basis. The calibrated model of Beijing shows that the coordination of local amenities provision has significant impacts on the performance of urban spatial strategies. Uncoordinated local amenities provision may undermine or even overturn the long-term plans for building a polycentric city region.

L. Wan (Corresponding author) • X. Rong • Y. Jin

Department of Architecture, University of Cambridge, UK

Email: lw423@cam.ac.uk

1. Introduction

The significance of the municipal amenities or public goods made available locally ('local amenities' below) to the choices made by the citizens has been well understood since the work of Tiebout (1956). Local amenities such as schools, hospitals and parks figure prominently when citizens choose where to live or work. The discussions on local amenities have gained new topicality as a result of the new focus upon the governance and subsidiarity of local funding decisions. Nevertheless, local amenities have rarely been considered in practical, policy-oriented spatial equilibrium models of residential and employment choice. Theoretical models that consider local amenities in the choice models show that they potentially have large impacts upon the results. In particular, the economies of scale in local amenities can foster further concentration from the city region to local neighbourhood levels, which in turn could influence the urban spatial pattern. Such theoretical insights have rarely been tested in applied models. The lack of empirical data has hampered both the verification of theoretical insights and practical policy applications. More specifically, the empirical analyses of local amenities require data with high geographical resolution, which were not available in the past except under rare circumstances.

However, the recent emergence of online data sources has opened up new opportunities. The online data is often provided for administrative purposes such as enabling the citizens to access the local services more easily and to report feedbacks. The interactions over the years between data providers and the citizens tend to improve the fidelity and currency of the datasets. In our experience a large proportion of the online administrative data is geo-coded. In addition, the online map services have added an increasing range of Point of Interest (POI) data. Often the online POI data and administrative data complement each other in data verification, e.g. the two type of sources can be fused together to achieve a greater level of accuracy even when using newly published data sources.

In this paper, we develop a new spatial equilibrium model for residential and employment location choice with an explicit representation of local amenities, responding to the availability of the newly emerged online data sources. Beijing seems to be a particularly good and typical case study for this topic for three reasons. First, like a large number of cities in emerging economies, the city is experiencing fast urbanization where the official spatial development aims for balanced, polycentric growth whilst the businesses and citizens tend to favour concentric, pancake-like expansion; the unequal spread of local amenities has been considered an important barrier to achieving the policy aims. Secondly, the fast urban growth has been accompanied by widening inequality both across the urban areas and among socioeconomic groups, where the access to local amenities is a prominent issue. Thirdly, in the last few years the online administrative and mapping data has grown rapidly, which provides new empirical data for modelling. We hope that our model will shed new light on the policy predicaments by analyzing the combined impacts of urban land-use development, transport investment and local amenities provision measures.

Section 2 below reviews the literature regarding the modelling of local amenities, and Section 3 sets out the main model equations. Section 4 and 5 respectively present the data and model application based on information collected from the city of Beijing. Section 6 concludes with considerations for further work.

2. Literature review

The term of local amenities or local public goods was first introduced into the spatial economic literature by Tiebout (1956). It suggested that as the local amenities are provided by local expenditure, they engender spatial competition among different localities through mobility of residents and jobs. As a rule, if citizens can move from one location to another with little friction, they will move to the location in which the mixture of services and taxes that provides them with the greatest net benefit. Tiebout's article has generated an enormous and continuing literature, much of it dealing with the identification of optimum provision of local amenities from a welfare perspective. Another major research agenda is the measurement of real estate capitalization for non-market local public goods using hedonic methods (Rosen, 2002, Berger et al., 2008). For instance Zheng and Kahn (2008, 2013) empirically examined the determinants of the pricing of new residential buildings as a function of physical attributes and access to various local public goods. They suggest that local public goods are important determinants of real estate prices, hence the choices of residential and job locations in Beijing.

The impacts of local public goods on urban spatial structure were first discussed by Stiglitz (1974), which explained how economies of scale in local public goods foster urban concentration. Helpman and Pines (1980) treated a system of monocentric cities without congestion, and found that, under free population mobility among the cities, the cities with high local amenities levels – if developed – will invest more in local public goods, be more populated and of higher population density. Although the spatial impacts of local amenities policies have been widely recognized since then, spatial modelling research with explicit representation of local amenities is rather limited. This is due to the fact that data on local amenities, particularly those with high spatial resolution, are difficult to collect from conventional data sources.

Since in an urban context, local amenities are intricately related to all aspects of production as well as those of consumption in space, spatial general equilibrium models are particularly appropriate tools for understanding the full implications of alternative strategies. For instance, a general equilibrium model can account for the sources of finance for the provision of local amenities, e.g. whether the level of finance is affordable given the levels of production and rental receipts without creating a large deficit. Also, the cost-minimizing behavior of businesses and utility-maximizing one for residents in spatial equilibrium models provide a solid basis for incorporating the effects of local amenities in location choice modelling.

Most existing spatial general equilibrium models cope with the lack of empirical data by adopting an egalitarian distribution of government revenue in the form of the rent dividend in a general sense. This evenly distributed rent dividend implies a uniform level of public investment shared by all citizens. Anas and Pines (2013) reviewed this strand of modelling practice, and provided a generic framework for studying the spatial impacts of local amenities and congestion in a system of equilibrium cities. In their model the cost of local amenities is fixed from demand side and independent of local population. The total expenditure is funded inside each city as endogenous optima with local taxation on land development and other fiscal instruments, such as congestion tolls. The monetary benefit of such local amenities is the costless commuting service for the urban central area, while there is no explicit local amenities utility measurement in the locational choice models.

This paper differs from the previous studies in three respects. First, instead of modelling the aggregate expenditure constraint on public goods as an endogenous optimum, our model deals with such constraints as exogenous policy inputs for each model period. This is a deviation from the general equilibrium framework (e.g. of Anas and Liu, 2007; Anas and Pines, 2013)

– In our model the inputs of the aggregate expenditure on local amenities is updated for each model period in a recursive manner, subject to background trends and policy targets under a recursive spatial equilibrium framework (Jin et al., 2013). This is dictated by the policy context of a market-oriented planned economy. Secondly, the proposed utility measurement from local amenities is directly supported by the emerging new data source at a relatively detailed zonal level, which enters the logit-type location choice model and is compatible with the spatial equilibrium structure. Thirdly, the model is calibrated to the real case of Beijing therefore the spatial impacts of local amenities can be tested and discussed with policy makers in practice. The integrated model framework is also applicable to other cities with similar development profile as Beijing.

3. Model design

The overall structure of the new model follows the shared convention between recursive and general spatial equilibrium models, i.e. the trade in labour, goods and services between locations is modelled simultaneously with the locational choices of production, consumption, jobs and homes. In the recursive spatial equilibrium model, the prices are determined at market equilibrium subject to recursively modified input constraints for each model period. The key equations of the model, which are introduced below, describe the behaviours of producers and consumers. In this context, we present the new model specifications for incorporating local amenities.

3.1. Producers

The producers are represented by a set of production functions that defines how they use capital, labour and properties. A Cobb-Douglas function with nested Constant Elasticity of Substitution (CES) has been broadly accepted as a standard for this purpose in spatial general equilibrium analysis since Krugman (1991) and Fujita et al (2001).

$$X_j^r = E_j^r A_j^r (K_j^r)^{v^r} \left(\sum_w (L_j^w)^{\frac{\theta^r}{\theta^r - 1}} \right)^{\frac{\delta^r (\theta^r - 1)}{\theta^r}} \left(\sum_k (B_j^k)^{\frac{\zeta^r}{\zeta^r - 1}} \right)^{\frac{\mu^r (\zeta^r - 1)}{\zeta^r}} \tag{3.1}$$

where X_j^r is the output of industry r in zone j. The main inputs to production are capital K, labour L, and building floorspace B. Constant internal returns to scale is assumed, where $v^r + \delta^r + \mu^r = 1$. For w varieties of labour and k varieties of building floorspace, a nested CES function is used to represent the substitution effects within each input bundle, the elasticity of substitution being governed by θ^r and ζ^r. A_j^r is a function of the economic mass for industry r in zone j that represents Hicksian-neutral Total Factor Productivity (TFP) effects resulting from learning and transfer of tacit knowledge (Rice et al., 2006, Graham and Kim, 2008), which are important component of urban agglomeration effects. Finally E_j^r is a constant scalar representing any additional zonal effects on total factor productivity, which is to be calibrated empirically.

Each type of building stock at zonal level is fixed for the model period as exogenous constraints, and can be updated periodically subject to background trends and planning targets. We follow standard assumptions that producers minimize the cost under budget and input supply constraints. The price of goods or services can then be derived as the average and marginal cost of production.

3.2. Consumers

For the consumers, we model how households source goods and services, their residence-employment location choice, and, for working households, determine how to divide time between work and leisure on the basis of utility and prices. Households are assumed to maximize utility under constraints of income and time. The utility measurement includes households' consumption of leisure time as well as goods and services, and housing.

(3.2)

$$V_{ij}^m = \alpha^m \ln[\sum_r (z_i^{mr})^{\frac{\eta^m}{\eta^m - 1}}]^{\frac{\eta^m - 1}{\eta^m}} + \beta^m \ln[\sum_k (b_{ij}^{mk})^{\frac{\sigma^m}{\sigma^m - 1}}]^{\frac{\sigma^m - 1}{\sigma^m}} + \gamma^m \ln(l_{ij}^w)$$

where V_{ij}^m define the consumption utility of household type m living in zone i and work in zone j; z_i^m is the final demand per household m for goods/services of type r from zone i; b_i^{mk} is the demand for housing type k per household m in zone i; l_{ij}^{w} is the leisure time in hours for employed worker w living in zone i and work in zone j. $\alpha^m + \beta^m + \gamma^m = 1$ are the shares of disposable income spent on goods/services, housing and leisure time, respectively. The elasticity of substitution is governed by η^m and σ^m between any two consumption varieties in goods/services and housing, respectively. Households may trade off consumption again leisure time and the unit value of time is measured by the hourly wage during working days of the year.

3.3. Local amenities and their financing

Local amenities are public goods that can be accessed only by residents in the local community: local schools, hospitals, parks, etc. are typical examples of local amenities. Local amenities exhibit non-excludability but are subject to service capacity and quality. Here we propose a log-form function to measure the extra utility gain from local amenities. The reason for choosing the log-form function is to facilitate the interaction between the local amenities utility component and the log-form household indirect utility function in the logit location choice model.

$$UA_i^m = f_i^m \sum_h \ln\left(1 + \frac{\tau_i^h N_i^h}{\sum_m H_i^m} \middle/ \frac{\sum_i \tau_i^h N_i^h}{\sum_{m,i} H_i^m}\right) \tag{3.3}$$

where $UA_i^m \geq 0$ is the total local amenities utility for household type m living in zone i; scalar parameter $f_i^m \in [0,1]$ measures the sensibility variance among different types of households; τ_i^h is a set of coefficients representing the service quality of local amenities type h in zone i, which is calibrated empirically; H_i^m is the number of household type m in zone i; finally $N_i^h \geq 0$ is the number of local amenities type h in zone i. Here we use the zonal number of key facilities, assuming that the number of key facilities is more perceivable to the local residents thus more representative to measure the utility gain. The function implies that the utility from local amenities is relative to the regional average level, and, for any given number of local amenities, the utility would decrease as the number of local users increases.

The utility from local amenities exerts impacts on the households' location choices and the log-form function facilitates a relatively easy interface with the existing spatial interaction models. Following the random utility interpretation of such models (McFadden, 1973), we define the location utility for household type *m* living in zone *i* and work in zone *j* as $v_{ij}^{m} = V_{ij}^{m} - d_{ij}^{m} + UA_{i}^{m} + \mathrm{E}_{ij}^{m} + \varepsilon_{ij}^{m}$, where V_{ij}^{m} is the consumption utility; d_{ij}^{m} is the generalized transport cost for household m commuting between zone *i* and *j*; $UA_{i}^{m} \geq 0$ is the local amenities utility; $\mathrm{E}_{ij}^{m} \in \left(-\infty, +\infty\right)$ is a constant term measuring the inherent attractiveness of discrete choice of living in zone *i* and working in zone *j*, which is internally calibrated; finally $\varepsilon_{ij}^{m} \sim$ *i.i.d. Gumbel* is a constant representing unobservable idiosyncratic variations in utility. This gives rise to the well-known multinomial logit choice probabilities:

$$P_{ij}^{m} = \frac{S_i \, e^{\lambda^{m}(V_{ij}^{m} - d_{ij}^{m} + UA_{i}^{m} + \mathrm{E}_{ij}^{m})}}{\sum_{i,j} (S_i \, e^{\lambda^{m}(V_{ij}^{m} - d_{ij}^{m} + UA_{i}^{m} + \mathrm{E}_{ij}^{m})})} \tag{3.4}$$

where P_{ij}^{m} is the probability of employed household type *m* choosing to live in zone *i* and work in zone *j*; S_i is a size term that corrects for the bias introduced by the uneven sizes of zones in the model (Ben-Akiva and Lerman, 1985); and λ^{m} is a dispersion parameter which is empirically calibrated.

Another important component associated with local amenities is government financing. The government financing in Chinese cities is complicated, as governments at different administrative levels have different taxation categories and retention shares. Although local public expenditure is mainly fund by district/county government, a certain proportion comes from city government. The amount of the allocated subsidy varies among districts/counties and is not readily traceable. All of these make it difficult to model the government financing in terms of individual funding streams. On the other hand, the published expenditure data on public services in Beijing is at the city district/county level, which is the total sum expenditure on local amenities from all financing sources.

Given the situation of data availability, we specify the modelling of government expenditure as follows. Following the tradition of spatial general equilibrium model in Anas and Pine (2013) we assume that the implicit government imposes a full Henry George tax on all property rents as the only source of government revenue. We then divide the Henry George tax based government expenditure into two parts: the first part is the observed expenditure on local amenities, where we take the observed expenditure by district and distribute it to model zones through the detailed online reports by facility (see Section 4); the second part is what is left of the Henry George tax based revenue after netting the local amenities expenditure, which is distributed equally per head of the residents, again following Anas and Pine (2013).

(3.5)

$$I_i^h = \ell_d^h \left(\sum_{k,i} B_j^k R_j^k \sum_{l,j} b_i^l r_i^l \right) \frac{\tau_i^h N_i^h}{\sum_{i \in d} \tau_i^h N_i^h}$$

where ℓ_d^h is the share of total rent dividend spent on providing local amenities type h in district d, which is exogenously observed and can be updated subject to policy scenarios; B_j^k and b_i^l are the stock sizes of business floorspace type k and housing floorspace type l, respectively; accordingly R_j^k and r_i^l are the business and housing floorspace rents per floorspace type per square meter. Due to the property of general equilibrium structure, this public investment needs to be transferred into monetary benefit that is evenly shared by the local residents.

(3.6)

$$MA_i = \sum_h \left(\frac{I_i^h}{\sum_m H_i^m} \right)$$

where MA_i^m is the public investment received by households living in zone i. The monetary gain from local amenities becomes part of the modelled income for the households, which in turn generates demand for goods and services in the spatial equilibrium framework.

3.4. Model outputs for policy assessment

The model outputs are quantities (production, factor inputs, and consumption demand) and prices, including wages and rents, in each zone, and the transport flows of people and goods/services between zones. Following the traditions of spatial econometric modelling we use the overall consumer surplus to measure the household well-being, which is defines as the change in average household utility divided by the average marginal utility of money:

$$\Delta C = \frac{(\bar{V}^{Alternative} - \bar{V}^{Base})}{\frac{1}{2}(\frac{1}{\bar{\Omega}^{Alternative}} + \frac{1}{\bar{\Omega}^{Base}})} \tag{3.7}$$

where $\bar{V}^{Base}$ and $\bar{V}^{Alternative}$ are the average household utilities; $\bar{\Omega}^{Base}$ and $\bar{\Omega}^{Alternative}$ are the average household incomes for the Base and Alternative scenarios, respectively.

4. Case study data

The model starts with inputs of transport supply and zonal stock of housing and business floorspace. For models of one city region, the total population size in the study area is also exogenously given, since there is no scope for modelling migration between city regions. For model calibration in the base year, it is necessary to input also the observed number of households and jobs and the expenditure on local amenities. To address the policy questions of social inequality, the model segments the households and employed workers into different socioeconomic groups.

4.1. Population and socioeconomic segmentation

For modelling population and employed workers we assume full employment, which is close to the situation in Beijing. The total number of employed residents in 2010 is derived by comparing the number of jobs and employed residents in each sector from several statistic sources[1]. The

---

[1] Data sources include Beijing Statistical Yearbook, China Statistical Yearbook, 2010 Population Census and 2008 Economic Census.

number of employed residents by work zone and that by residence zone are based on the neighbourhood (*Jiedao)*[2] level employment data in the 2008 Economic Census and the 2010 Population Census.

Based on the well-established Erikson-Goldthorpe-Portecarero (EGP) schema (Goldthorpe et al., 1980), the model differentiates the employed residents into three socioeconomic groups, which as a shorthand reflect broadly the income levels of the respective groups. More specifically, the model uses the occupation statistics in Population Census and the categorization method used by Chen (2013) and Treiman (2012) to calculate the total number of employed residents in each socioeconomic group. The zonal level data is then calculated based on education level data of employed workers and residents in 2008 Economic Census and 2010 Population Census. The employed population is summarized in Table 1.

Table 1. Data summary - employed population in Beijing 2010

	High-income	Middle-income	Low-income
Centre	882931	2501870	858901
	47%	37%	26%
Near suburbs	558451	1713386	695128
	30%	26%	21%
Far suburbs	226784	1175938	759257
	12%	18%	23%
New town	165248	1045333	669489
	9%	16%	21%
Ecological protection area	32522	241963	278354
	2%	4%	9%
Sum	1865936	6678490	3261129
	100%	100%	100%

[2] *Jiedao* is the smallest administrative unit in China and the finest geographic level for government statistics.

4.2. Building stock and generalized travel costs

We use building floorspace area to measure the size of the building stock in Beijing at the zonal level. Due to the difficulties in accessing the zonal data for business floorspace, we derive the zonal business floorspace by assuming that each employed person uses 20 m^2. This is a crude assumption, although it is broadly fits the data currently available and can be further refined by business sector when required. Zonal housing floor space is calculated based on number of residents and average housing floorspace per capita in 2010 Population Census at district level. The building floorspace data is summarized in Table 2.

Table 2. Data summary - building stock in Beijing 2010

(million m^2)	**Business floorspace**	**Housing floorspace**
Centre	118.4	184.3
	50%	32%
Near suburbs	41.0	139.5
	17%	24%
Far suburbs	32.9	116.0
	14%	20%
New town	35.4	104.9
	15%	18%
Ecological protection area	9.2	30.7
	4%	5%
Sum	236.8	575.4
	100%	100%

The generalized travel costs between any zone pair is computed from the monetary cost and travel times from a multimodal transport model for Beijing. The travel times include traffic congestion as derived from observed road travel data from 2008.

4.3. New data sources for modelling local amenities

Modelling local amenities at zonal level requires local amenities data at a high spatial resolution that conventional published sources can hardly support. However, with the emerging of electronic maps and the Open Government scheme in China, new data sources open up new possibilities for local amenities modelling.

The first of the new data sources is the GIS-based Point of Interest (POI) provided by major electronic map services such as Google, Sogou and Baidu[3]. The POI data are geo-coded and include information for various kinds of local facilities, such as grocery stores, hospitals and post offices. Our own checking using local knowledge suggests that the POI information is broadly accurate, having been in use by online browsing for a number of years.

The second new data source is the E-maps initiated by the municipal government of Beijing for promoting the Open Government scheme. The E-map data source shares the same advantages of POI data, as it is usually jointly produced by the government and the map service providers. More importantly, the official resources available for their production and the fact that they have been used on a day to day basis for a number of years has meant that their quality and coverage are good in most cases.

Based on the range of data available, we focus on two types of local amenities: state-funded[4] schools (both primary and secondary schools[5]) and hospitals. There are three reasons for this choice. First, schools and hospitals are typical public facilities in Beijing and their significances in residence location choice and hedonic housing price have been widely recognized (Hu et al., 2014, Zheng and Kahn, 2008). Secondly, the public expenditure

[3] Sogou (http://map.sogou.com/) and Baidu (http://map.baidu.com/) are both leading online map service suppliers in China.
[4] We do not include private/international schools for two reasons: (1) the data on private/international schools are not available in official statistics therefore their number, rating and financing process are not clear; (2) most of these schools are boarding/lodge schools which receive students citywide, thus are not typical local amenities by definition.
[5] We do not include high schools because they recruit students citywide, whereas primary schools and secondary schools recruit students mostly within the local area. Secondary schools can recruit some students from outside the local area, but the percentage is usually lower than 50% (Hu et al., 2014). More specifically the data on secondary schools include “complete schools” which have both secondary and high schools, and also “nine-year schools” which have both primary and secondary schools. By definition the “nine-year schools” are also listed in the primary schools. We allow duplication in counting as it implicitly reflects the size of the schools.

data on education and public health at the district level are accessible as existing categories in the government statistics. The corresponding data at aggregate level enables the cross-check of the new data source. Thirdly, the quality categorizations of schools and hospitals are issued according to official standards and are included in the new data sources, which avoid arbitrary rating of quality difference.

For schools, we use the Beijing Education Map from the official website of Beijing Municipal Commission for Education[6], which contains 1825 primary and secondary schools in total. For hospitals, which do not yet have a government E-map, we collect the POI data from Sogou online map. The collected POI data includes information on quality rating and has a good coverage (5274 hospitals of all official ratings).

Given the zoning configuration of the modeled area, geo-processing methods are then applied to calculate the zonal number of schools/ hospitals of different quality categories. The zonal total number of facilities per type is used to subdivide the observed public expenditure from district level to zonal level. In addition to the quality difference between categories, we find that the average local amenities level could also vary among districts. For instance two hospitals in different districts, though entitled with the same quality rating, may have distinct service level and reputation. According to our local experience this locational difference can be well perceived by residents. In order to address this unobservable aspect of quality difference, we define a set of locational quality coefficients by calculating the relative investment intensity per facility.

$$\tau_d^h = \frac{(\ell_d^h / \sum_{i \in d} N_i^h)}{\sum_d \ell_d^h / \sum_i N_i^h} \tag{4.1}$$

where τ_d^h is the zonal quality coefficient for local facility type h in district d; ℓ_d^h is the public expenditure on local facility type h in district d; N_i^h is the number of facility type h in zone i, which is obtained from the new data sources. The quality coefficients are uniform for all zones within the district and are used to adjust the number of facilities so that the quality difference among districts can be captured in the model. Table 3 provides a summary of adjusted number of schools and hospitals in Beijing by location and quality category.

[6] http://www.beijingmap.gov.cn/bjjw/

Table 3. Data summary – schools & hospitals in Beijing 2010

	Adjusted number of Schools		Adjusted number of Hospitals		Employed Population (million)
	All	Key	All	Key	
Centre	738	146	2379	189	5.06
	40%	71%	45%	60%	36%
Near suburbs	307	26	807	38	3.52
	17%	13%	15%	12%	25%
Far suburbs	361	9	815	18	2.55
	20%	4%	15%	6%	18%
New town	251	23	928	47	2.24
	14%	11%	18%	15%	16%
Ecological protection area	167	1	344	25	0.65
	9%	0%	7%	8%	5%
Sum	1825	205	5274	317	14.02
	100%	100%	100%	100%	100%

Table 3 clearly shows that public schools and hospitals are highly concentrated in central Beijing. More than 70% of the key schools and 60% of the key hospitals are located in the centre whilst the central population only accounts for 36% of the municipality.

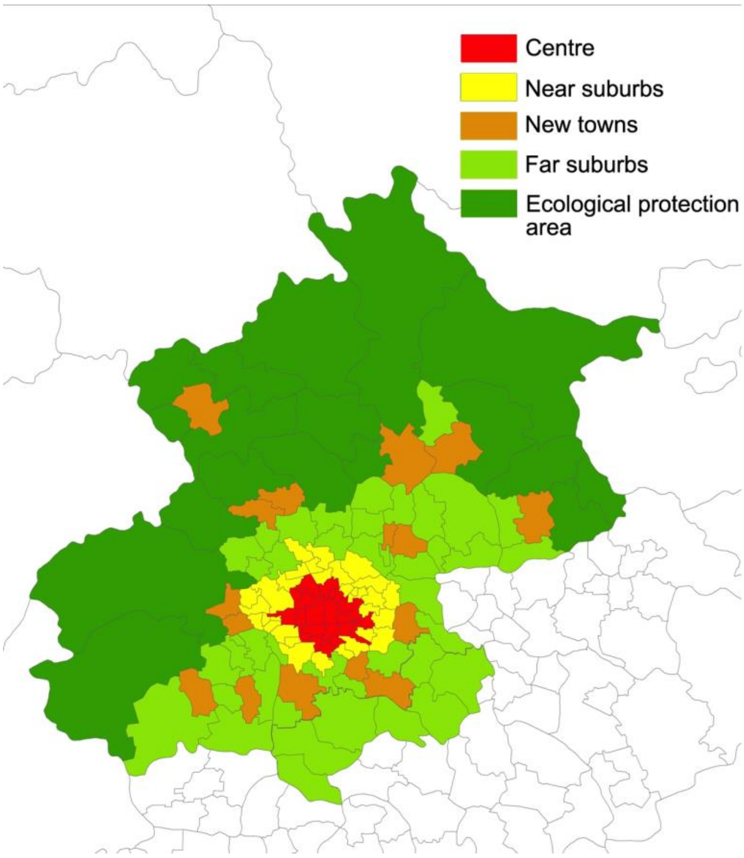
5. Model tests

The study area (i.e. Beijing municipality) is represented with 130 zones (Figure 1). The zoning is based on the administrative boundaries of *Jiedao*. The model follows the tradition of having smaller zones closer to the centre and larger zones in the peripheries which is in line with population density. In order to better model the policy scenarios, zones are categorized into five types by location: Centre, Near suburbs, New towns, Far suburbs and Ecological protection areas, according to the Municipal Master Plan of Beijing 2004-2020.

Table 4. Zone numbers by category

Area	Number of model zones
Centre	23
Near suburbs	31
New towns	16
Far suburbs	42
Ecological protection area	18
All	130

Figure 1. Model zoning by category



5.1. Model parameterization

We take parameter values from the data and established models, following Jin et al (2013). Where there are no commonly accepted parameters we carry out sensitivity tests in the model and adopt value ranges by judgment. Table 5 lists the model parameters that have been specified in the equations.

Table 5. Model parameters and the sources

Model parameter	Value(s)	Sources
ν^r (capital cost share)	0.00	Anas and Rhee (2006)
δ^r (labour cost share)	0.76	Beijing statistic data (2010)
μ^r (business floorspace share)	0.24	Beijing statistic data (2010)
E_j^r (total factor productivity multiplier)	1.00	Anas and Rhee (2006)
θ^r (elasticity for labour variety)	0.80	Own sensitivity tests
α^m (household utility parameter for goods/service)	0.37* 0.375 0.365	Beijing statistic data (2010)
β^m (household utility parameter for housing)	0.13* 0.125 0.135	Beijing statistic data (2010)
γ^m (household utility parameter for leisure time)	0.50* 0.50 0.50	Own sensitivity tests
λ^m (dispersion parameter for spatial interaction model)	1.0 1.1 1.2	Calibrated to reproduce an average commuting distance that is compatible with 2010 Beijing travel data
f^m (household sensitivity parameter for local amenities)	0.8 0.5 0.1	Calibrated to reproduce zonal number of households that is compatible with 2010 Beijing data
Total number of working days per year	250	Anas and Rhee (2006)
Hours per day	24	Anas and Rhee (2006)

* Row values are for high-income, middle-income and low-income group, respectively.

5.2. Model runs

We present three types of model runs to highlight the key features of the new model component (See Table 6). The forecast runs for 2010 in Test 1 are counterfactual and are conducted with no inherent residual attractiveness parameters (E_{ij}^{m}). The purpose of this test is to measure to what extent the new local amenities component could improve the model's goodness of fit. The calibration runs in Test 2 follow the standard calibration procedure for spatial equilibrium models, where the E_{ij}^{m} parameters are incorporated and solved through an optimization algorithm subject to observed location pattern of homes and jobs. Test 2 aims to measure the percentage share of the local amenities element in the unobserved residual attractiveness parameters. Test 3 is a set of static equilibrium runs for 2050 with updated constraints and boundary conditions that reflect the broad long-term trends of development in Beijing – they are designed to test the capabilities of the new local amenities component in assessing new policy scenarios.

Table 6. Summary of model runs

Test	**Year**	**Run Mode**	**New Local amenities Component (MA_i , UA_i^m)**	**Inherent attractiveness (E_{ij}^m)**
1	2010	Counterfactual forecasts	√	×
	2010	Counterfactual forecasts	×	×
2	2010	Calibration	√	√
	2010	Calibration	×	√
3	2050	Scenario forecasts	√	√*

* Parameters are inherited from the calibration runs for 2010 and remain constant in all scenario forecasts

5.2.1. Counterfactual forecasts for 2010

In this test the model starts with the same inputs of generalized travel costs, housing and business floorspace stock, and the municipal totals of households of each socioeconomic group. We compare the goodness of fit of the modeled zonal number of residents with and without the local amenities component. The goodness of fit is measured by a weighted sum of squared errors:

$$e_m = \sum_i \frac{(H_i^{m|\mathrm{mod}} - H_i^{m|obs})^2}{H_i^{m|obs}} \tag{5.1}$$

where $H_i^{m|\mathrm{mod}}$ is the modeled zonal number of household type m in zone i and $H_i^{m|obs}$ is the observed number correspondingly. In order to compare the change of sum error, we further define the percentage of sum error change as $(e^{\surd} - e^{\times})/e^{\times} \times 100\%$, where $e^{\surd}$ and $e^{\times}$ is the sum error with and without the local amenities component, respectively.

Table 7. Goodness of fit with & without the local amenities component

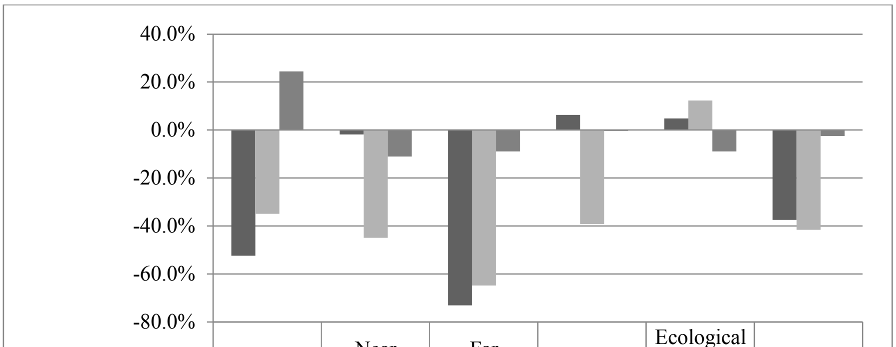
Household types	Local amenities component	Linear regression function	R^2	Sum error (e_m)	% of error change
High income (f = 0.8)	×	y = 0.3591x + 9199	0.7381	1.39E+06	-
	√	y = 0.6866x + 4498	0.7624	8.68E+05	-37.5
Middle income (f = 0.5)	×	y = 0.5161x + 24861	0.8494	1.36E+06	-
	√	y = 0.7778x + 11414	0.8574	7.93E+05	-41.7
Low income (f = 0.1)	×	y = 0.5042x + 12438	0.6153	7.72E+05	-
	√	y = 0.5270x + 11865	0.5981	7.53E+05	-2.5
Overall	×	y = 0.5417x +	0.8654	3.52E+06	-

		41623			
	√	y = 0.7597x + 21826	0.8414	2.41E+06	-31.4

Overall the incorporation of local amenities brings a decrease of sum error by 31.4%. The distinct responses of different household types indicate that households of different social-economic backgrounds have different degrees of sensitivity towards local amenities. Significant improvements are witnessed for the high-/middle-income groups. In contrast the minor improvement (2.5%) for the low-income group implies that local amenities level is likely to be a less effective factor in location choices of this group.

Figure 2 shows the breakdown of sum error reduction per household type by location. In near suburbs and far suburbs, the incorporation of local amenities reduces the sum error for all household types. However exceptions of reverse effect occur to the high-/low-income groups in certain locations. Specifically the new local amenities component exaggerates the attractiveness in new towns for the high-income group, which results in a further increase of sum error. This also explains the overestimation of low-income households in the centre. In the ecological protection area, the numbers of high-/middle-income groups are overestimated; whilst the sum error for the low-income group is slight decreased.

Figure 2. Percentage change of sum error by location per household type



	Centre	Near suburbs	Far suburbs	New town	Ecological protection area	Overall
High_income	-52.5%	-1.9%	-73.1%	6.3%	4.8%	-37.5%
Middle_income	-35.0%	-45.0%	-64.8%	-39.3%	12.3%	-41.7%
Low_income	24.5%	-11.0%	-8.9%	-0.4%	-8.8%	-2.5%

5.2.2. Calibration runs for 2010

The calibration runs differ from the counterfactual forecasts in the incorporation of the inherent attractiveness parameters (E_{ij}^{m}), which represent the inherent utility for choosing the discrete residence-employment location pair (*i, j*) for household type *m*. In calibration runs E_{ij}^{m} is calibrated through iterative algorithm according to the targeted commuting OD data. In our model the E_{ij}^{m} parameters enter the location choice model as quantities by zone pair and the calibration process enables the model to reproduce of the targeted OD for each household type. In order to measure the inherent attractiveness for only residence zones, rather than zone pairs, we further introduce the E_{i}^{m} parameter, which is the log-sum aggregation of the E_{ij}^{m} matrix.

(5.2)

$$\mathrm{E}_{i}^{m} = \frac{1}{\lambda^{m}} \log\left(\sum_{j} e^{\lambda^{m}(\mathrm{E}_{ij}^{m} - \bar{\mathrm{E}}^{m})}\right)$$

where E_{i}^{m} represents the aggregated unobservable attractiveness of residential zone *i* for household type *m*, and $\bar{\mathrm{E}}^{m}$ is the numeric average of the E_{ij}^{m} matrix. The reason for deducting the average is to make the values of E_{i}^{m} numerically comparable among different model runs[7]. By comparing the change of the aggregated zonal parameters E_{i}^{m} with and without the new local amenities component, we can estimate the percentage share of local amenities in the overall residual utility.

(5.3)

$$k_{i}^{m} = \frac{UA_{i}^{m}}{\mathrm{E}_{i}^{m}} \times 100\% \quad \text{subject to } \mathrm{E}_{i}^{m} = \mathrm{E}_{i}^{*m} + UA_{i}^{m}$$

where UA_{i}^{m} is the extra utility gain from local amenities in zone *i* for household type *m*; E_{i}^{m} is the calibrated aggregate attractiveness for zone *i*

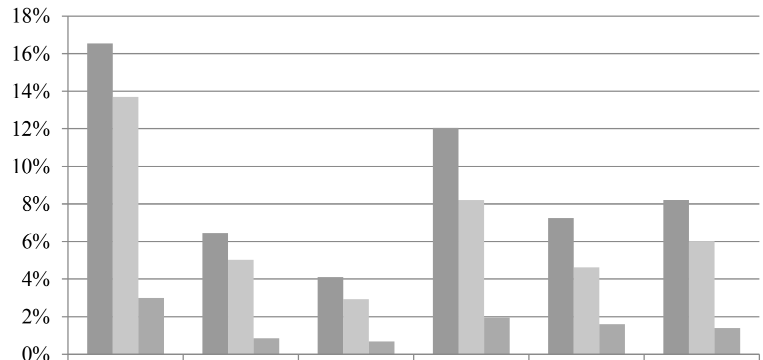
[7] Note that deducting the $\bar{\mathrm{E}}^{m}$ variable only changes the average value of the log-sum function. It does not change the probability distribution of Equation 3.4, which ensures that the model is calibrated to the same equilibrium status.

without the local amenities component; E_i^{*m} is the calibrated attractiveness for zone i after adding the local amenities component into the logit model. The values of E_i^m, UA_i^m and k_i^m are presented in Table 8 and Figure 3.

Table 8. Values of $\mathbf{E}_i^m$ and $\boldsymbol{UA}_i^m$ per household type by location

	$\boldsymbol{E}_i^m$				$\boldsymbol{UA}_i^m$			
	High-inc.	Mid-inc.	Low-inc.	Ave.	High-inc.	Mid-inc.	Low-inc.	Ave.
Centre	6.3	4.7	4.5	5.2	1.0	0.6	0.1	0.6
Near suburbs	6.7	5.2	4.7	5.6	0.4	0.3	0.0	0.2
Far suburbs	6.4	5.6	4.7	5.5	0.3	0.2	0.0	0.2
New town	6.2	5.6	4.7	5.5	0.7	0.5	0.1	0.4
Ecological protection area	6.0	6.1	4.8	5.6	0.4	0.3	0.1	0.3

Figure 3. Percentage share of the local amenities component in the overall residual utility



	Centre	Near suburbs	Far suburbs	New town	Ecological protection area	Overall
High_income	17%	6%	4%	12%	7%	8%
Middle_income	14%	5%	3%	8%	5%	6%
Low_income	3%	1%	1%	2%	2%	1%

The existence of E_i^m parameters indicates, unsurprisingly, that the mere consideration of travel cost and building stock supply cannot reproduce the observed zonal number of residents. Table 8 shows that model without E_i^m parameters underestimates the number of high-income households in near suburbs, and also the number of middle-income group in far suburbs and new towns. Figure 3 further shows that, for high-income households, the local amenities component UA_i^m explains up to 17% of the total unobservable attractiveness in the center. This utility percentage drops for the lower income groups. The percentage also differs by location - Generally it is higher in the centre and new towns, where the average local amenities levels are also higher than other locations.

5.2.3. Scenario tests for 2050

In this section we further run the model for 2050 to demonstrate how different local amenities provision policies can interact with land-use and transport supply, and influence the outcome of decentralization strategy in Beijing. The forecast model for 2050 is based on the calibrated 2010 model, with the new local amenities component incorporated.

In terms of boundary conditions we assume a high-growth scheme[8] where the population will double to 40 million following Wu (2012), among which 50% are employed residents, compared to 52.6% in 2000 and 60.3% in 2010. The decreased employment rate reflects fact of aging population and extended study/training periods in line with the developed countries today. For average income, we follow OECD's optimistic projection of China's GDP per capita which implies an increase of average income by 7.3 times from the current level – the Beijing city region is likely to attain this level of growth given its national capital status. In terms of socio-economic composition, we assume that China will follow the EU model of larger middle-income groups and smaller low-income groups. We also assume that business floor space and housing floor space grow at the same pace as the number of employed residents and number of total population,

[8] Other development schemes such as population control scheme can also be applied. However, because the local amenities in our model is measured by per capita value and the scenario settings are also based on per capita investment and endowment, changing the aggregate population level would not damage the generali-generality of the model results.

respectively. The projections of the socio-demographic and economic conditions in 2050 are summarized in Table 9 and Table 10.

Table 9. Demographic and economic projection in 2050

Year	**2010**	**2050**
Population	19,578,961	40,000,000
Employed population (Employment rate)	11,805,556 (60.3%)	20,000,000 (50.0%)
Average Income per capita (Yuan)	22,246	161,766

Table 10. Percentage share of socio-economic groups in 2050

Socio-economic group	**2010**	**2050**
High-income	16%	25%
Middle-income	56%	65%
Low-income	28%	10%

The design of the 2050 scenarios is broadly based on the government's policy objective to establish sub-centres in new towns through housing and employment decentralization (BMG, 2005). Particularly housing is decentralizing at a faster speed than employment where offices and institutions in Beijing are still developing in a centralized manner between 2004 and 2013 (Rong et al., 2014). Based on this broad trend, we assume half of the zonal business and housing floorspace growth comes from natural growth, which is proportional to the existing stock size in each zone. And the other half is defined as discretionary growth, which is directed by the alternative spatial planning strategies.

The assumptions for the discretionary building stock growth reflect the continuing growth of employment in the centre and the fast catch-up in new towns. The discretionary component of business floorspace is allocated to the centre (60%) and the new towns (40%). To reflect the decentralization of housing and the policy orientation on Transport-Oriented-Development (TOD), the discretionary growth component for housing floorspace is allocated to the model zones in the new towns (70%) and far suburbs that have with metro stations (30%). Within each targeted area (i.e. the centre, new towns and far suburb zones with metro stations), the discretionary growth in floorspace is allocated to the component zones based on the existing stock size in each zone. The building floorspace supply assumptions for 2050 are summarized below.

Table 11. Building floorspace scenario 2010-2050

	2010 (million m^2)		**2050** (million m^2)		**Growth Rate 2010-2050**	
	Business	Housing	Business	Housing	Business	Housing
Centre	118.4	184.3	209.4	276.5	76.9%	50.0%
Near suburbs	41.0	139.5	55.1	209.2	34.6%	50.0%
Far suburbs	32.9	116.0	44.3	260.3	34.5%	124.4%
New town	35.4	104.9	80.7	358.8	127.9%	241.9%
Ecological protection area	9.2	30.7	11.9	46.0	29.7%	50.0%
Sum	236.8	575.4	401.4	1150.8	69.5%	100.0%

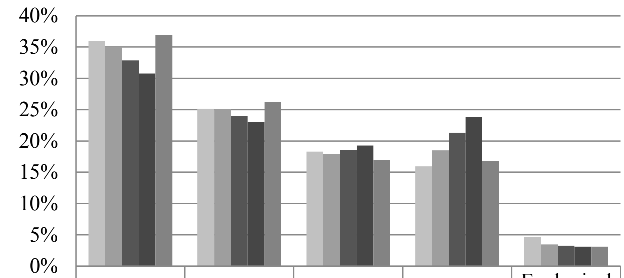
In order to highlight the key features of the local amenities component, we develop four alternative scenarios of local amenities investment: (1) *trend growth* where the provision of amenities in 2050 replicates the zonal patterns as observed in the early 2010s; (2) *coordinated growth* which allocate the investments in line with new housing growth; (3) *coordination-enhanced growth* where local amenities investment in new towns and far suburb towns with metro-station are further increased by 150% based on the coordinated growth scenario, while the regional sum remains the same; (4) *uncoordinated growth* where all local amenities development concentrates in the center and near suburbs. The specifications of the local amenities scenarios are summarized in Table 12.

Table 12. Local amenities scenarios in 2050

Number of zonal schools + hospitals	**2050 Scenarios**							
	Trend		Coordinated		Coordination Enhanced		Uncoordinated	
	Key	All	Key	All	Key	All	Key	All
Centre	706	5983	585	4676	455	3170	802	7919
Near suburbs	136	2136	112	1672	87	1138	155	2829
Far suburbs	56	2253	82	2468	110	2733	31	1177
New town	146	2266	274	4034	412	6050	80	1180
Ecological protection area	54	978	44	767	33	525	29	511
Sum	1097	13616	1097	13616	1097	13616	1097	13616

We use the predicted number of residents by location to demonstrate how different local amenities provisions can affect the performance of urban decentralization strategy. The results are summarized in Figure 4. In base year 2010, 61% of population lives in centre and near suburbs. In the trend scenario, 60% of the total population still lives in the centre and near suburbs, which essentially retains the centralized pattern from 2010. The coordination-enhanced scenario produces the highest degree of decentralization, with 43% of total population living in far suburbs and new towns. By contrast the decentralized population only accounts for 34% in the uncoordinated scenario. It indicates that the physical decentralization of business and housing floorspace may not necessarily lead to the polycentric pattern, if local amenities provision still concentrates in the city centre.

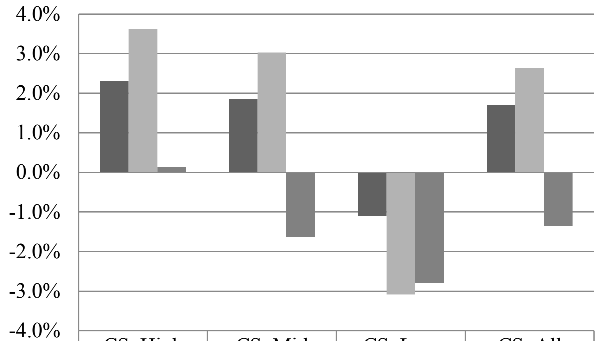
Figure 4. Population distribution in 2050 scenarios



	Centre	Near suburbs	Far suburbs	New town	Ecological protection area
2010_Baseline	36%	25%	18%	16%	5%
2050_Trend	35%	25%	18%	19%	3%
2050_Coordinated	33%	24%	19%	21%	3%
2050_Coordination_Enhanced	31%	23%	19%	24%	3%
2050_Uncoordinated	37%	26%	17%	17%	3%

In terms of economic welfare, Figure 5 shows the consumer surplus per household type and the overall value based on the 2050 trend growth scenario. The coordination-enhanced scenario again achieves the highest consumer surplus of 2.6%; while the uncoordinated provision of local amenities could reduce household welfare by an equivalent of 1.4% of average income. The decentralization gets the high-income group better off up to 3.6%, due to the fact that the increased commuting cost is compensated by larger housing space in suburbs. The low-income group however suffers from a global reduction of consumer surplus in all decentralization scenarios. This is mainly caused by: (1) as population gets decentralized there is increasing competition for housing floorspace in suburbs, resulting in higher rents and smaller housing space for the poor; (2) as the business floorspace and jobs still concentrate in the centre, higher transport costs are incurred by long commuting from the suburbs to the centre.

Figure 5. Consumer surplus in 2050 local amenities scenarios



	CS_High	CS_Mid	CS_Low	CS_All
2050_Coordinated	2.3%	1.9%	-1.1%	1.7%
2050_Coordination_Enhanced	3.6%	3.0%	-3.1%	2.6%
2050_UnCoordinated	0.1%	-1.6%	-2.8%	-1.4%

6. Conclusion

This paper presents a new spatial equilibrium model that accounts for the detailed patterns of local amenities investment. This is supported by the new online data on local amenities that has not been available until recently. The use of the new data sources can improve the model's overall goodness of fit by over 30%. The improved model calibration shows households of different socioeconomic background have varying degrees of sensitivity to the local amenities in their residence location choice. The degree of sensitivity increases with the raise of household income level.

Moreover the new local amenities component provides a new interface for spatial equilibrium models and opens up new areas for policy simulation, where the combined effects of urban land-use, transport and local amenities policies can be simulated on a consistent basis. The integrated model framework may help the initiatives to develop a new, polycentric city region to accommodate high-growth pressures and to achieve a more balanced, flexible and environmentally sustainable configuration of urban growth. Our preliminary tests demonstrate that the policy objective of building a polycentric Beijing may be substantially undermined with uncoordinated development of local amenities.

However, this is an initial attempt to apply the new model and data sources, and there is much room for improvement. Firstly, the financing of local amenities needs to take local funding mechanisms into consideration. Local funded amenities can trigger the competition among communities and enable the modeling of gentrification and social segregation. Secondly, the parameterization for the new local amenities component requires further sensitivity tests, and should be validated once time-series data are available. The temporal dynamics of the spatial impacts from local amenities policies should also be further investigated through updating the policy variables in a recursive manner.

References

Anas, A., & Liu, Y. (2007). A regional economy, land use, and transport model (RELU-TRAN©): formulation, algorithm design, and testing. *Journal of Regional Science, 47*(3), 415-455.

Anas, A., & Pines, D. (2013). Public goods and congestion in a system of cities: how do fiscal and zoning policies improve efficiency? *Journal of Economic Geography, 13*(4), 649-676.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9): MIT press.

Berger, M. C., Blomquist, G. C., & Peter, K. S. (2008). Compensating differentials in emerging labor and housing markets: Estimates of quality of life in Russian cities. *Journal of Urban Economics, 63*(1), 25-55.

Chen, M. (2013). Intergenerational Mobility in Contemporary China. *Chinese Sociological Review, 45*(4), 29-53.

Fujita, M., Krugman, P. R., & Venables, A. J. (2001). *The spatial economy: Cities, regions, and international trade*: MIT press.

Goldthorpe, J. H., Llewellyn, C., & Payne, C. (1980). *Social mobility and class structure in modern Britain*: Clarendon Press Oxford.

Graham, D. J., & Kim, H. Y. (2008). An empirical analytical framework for agglomeration economies. *The Annals of Regional Science, 42*(2), 267-289.

Helpman, E., & Pines, D. (1980). Optimal public investment and dispersion policy in a system of open cities. *The American Economic Review*, 507-514.

Hu, W., Zheng, S., & Wang, R. (2014). The capitalization of school quality in home value: a matching regression approach with housing price-rent comparison (Chinese). *China Economic Quarterly, 13*(3).

Jin, Y., Echenique, M., & Hargreaves, A. (2013). A recursive spatial equilibrium model for planning large-scale urban change. *Environment and Planning B: Planning and Design, 40*(6), 1027-1050.

Krugman, P. R. (1991). *Geography and trade*: MIT press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Rice, P., Venables, A. J., & Patacchini, E. (2006). Spatial determinants of productivity: analysis for the regions of Great Britain. *Regional Science and Urban Economics, 36*(6), 727-752.

Rong, X., Jin, Y., & Ying, L. (2014). *Understanding Beijing's Urban Land Use Development 2004-2013 through Online Administrative Data Sources*. Paper presented at the The 8th International Association for China Planning (IACP) Conference.

Rosen, S. (2002). Markets and diversity. *American Economic Review, 92*(1), 1-15.

Stanford University. Institute for Mathematical Studies in the Social Sciences, & Stiglitz, J. E. (1974). *The theory of local public goods*.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy*, 416-424.

Treiman, D. J., Lu, Y., & Qi, Y. (2012). New approaches to demographic data collection. *Chinese Sociological Review, 44*(3), 56-92.

Wu, L., & Wu, W. (2012). *Beijing 2049: study on spatial development strategies*: Tsinghua University Press.

Zheng, S., & Kahn, M. E. (2008). Land and residential property markets in a booming economy: New evidence from Beijing. *Journal of Urban Economics, 63*(2), 743-757.

Zheng, S., & Kahn, M. E. (2013). Does government investment in local public goods spur gentrification? Evidence from Beijing. *Real Estate Economics, 41*(1), 1-28.