The Birth of Edge Cities in China:

Measuring the Spillover Effects of Industrial Parks

Siqi Zheng[1], Weizeng Sun[1], Jianfeng Wu[2] and Matthew E. Kahn[3]

1: Hang Lung Center for Real Estate, Tsinghua University

2. School of Economics, Fudan University

3. UCLA and NBER

July 11th 2015

Abstract

Several Chinese cities have invested billions of dollars to construct new industrial parks. These place based investments solve the land assembly problem which allows many productive firms to co-locate close to each other. The resulting local economic growth creates new opportunities for real estate developers and retailers that develop properties and stores close to the new park. The city mayor has the political clout and the personal promotion incentives to anticipate these effects as he chooses whether and where within the city to build the park. Using several geo-coded data sets, we measure the localized spillover effects of the new parks on local incumbent firm productivity, the growth of retail activity close to the park and local real estate pricing and construction. We document the heterogeneous effects of investment in parks. Those parks featuring a higher level of human capital, a greater level of co-agglomeration among firms within the park, and a smaller share of State Owned Enterprises offer greater spillover effects.

We thank Nick Mueller and participants at the January 2015 ASSA, USC, LSE and University College, London and UNLV for useful comments. We thank the UCLA Rosalinde and Arthur Gilbert Program in Real Estate, Finance and Urban Economics for generous funding. Jianfeng Wu thanks School of Economics, Fudan University, for making this project possible through the access to the ASIF database of China's manufacturing firms; He also thanks the support from the MOE Project of Key Research Institute of Humanities and Social Sciences at China Center for Economic Studies (CCES), Fudan University, and Research Institute of Chinese Economy (RICE), Fudan University. We are indebted to Zhikuo Liu for helpful discussions. Siqi Zheng and Weizeng Sun thank the National Natural Science Foundation of China (No. 71273154 and No. 71322307) and the Program for New Century Excellent Talents in University (NCET-12-0313) for research support.

Introduction

The rise of edge cities in the United States has fostered urban growth through providing global corporations with the advantage of cheaper land, and advanced new capital infrastructure and a high quality of life for employees and executives (Garreau 1991). Over the last twenty years, China's local officials have made much larger place based investments as they have sought to create new suburban edge cities. These new industrial parks are contained geographic areas within a city with special land, tax, financial and economic policies to recruit highly productive manufacturing and service firms (Wang 2013). Physical proximity between firms who seek to co-agglomerate facilitates local productivity growth. The well paid workers at these new employment centers seek nearby housing and retail opportunities. Real estate developers and retail entrepreneurs respond to this increased local market potential by supplying new housing and consumer amenities. This "chain reaction" of new jobs causing real estate construction and the opening of new high end retail opportunities creates a vibrant edge city.

Through its multi-billion dollar place based investments, China's central and local governments have made a large bet that the agglomeration benefits of industrial parks offer a stream of benefits that exceed the upfront cost of establishing the parks and the opportunity cost of allowing the land that the park sits on to be allocated for other purposes. Local governments have paid for place based projects using debt financing. According to IMF estimates, local-government debt reached 36% of GDP in 2013, double its share of GDP in 2008, and will increase to 52% of GDP in 2019. The popular press has suggested that China's local governments face a potential major debt crisis.[1] This ugly scenario is less likely to occur if place based investments generate significant economic spillovers. Wang (2013) estimates that industrial parks in Chinese cities increase the total factor productivity, foreign direct investment (FDI) and local workers' earnings.

China's unique political system grants city mayors with powers that far exceed their Western counterparts. China's local leaders have strong incentives to pursue economic growth because this increases their promotion chances. New industrial parks

[1] See http://www.wsj.com/articles/debt-that-once-boosted-its-cities-now-burdens-china-1422415981

can stimulate economic growth because they solve a co-ordination problem and a land assembly problem (Combes et. al. 2012). Establishments that seek to co-agglomerate within a small geographic area face transaction costs in simultaneously seeking to co-locate. Even if they could agree to do so, there is no reason to believe that a major city would have a large enough vacant plot of land that would allow them to simultaneously execute this plan. In this sense, the industrial park permits a degree of co-ordination in a timely fashion that is unimaginable in a setting featuring a land market with pre-existing durable structures and long lags between sales dates.

This paper uses detailed within city geocoded data to study the Chinese local governments past success in creating vibrant new edge cities. Our metrics of "success" focus on economic dynamics within cities in geographic areas close to the new parks. We test whether successful "work-live-play" places emerge based on outcome criteria such as localized job growth, TFP increases for incumbent firms, new housing construction, increased home prices, and increased retail opportunities. These outcomes directly benefit the local government officials because their fiscal revenue is tied to the commercial and residential land sales, and also the tax from productive firms.

To test for the extent of localized agglomeration spillovers, we merge together several geocoded data sets in China's eight major cities – 120 industrial parks, the productivity and employment data of manufacturing firms between the years of 1998 and 2007, and the housing and retail data after 2006. This piece of our empirical work builds on research documenting spatial industrial spillover effects (Greenstone, Hornbeck, Moretti 2010, Arzaghi and Henderson 2010).

Previous research studying the effects of China's industrial parks has focused solely on the productivity impacts on the select set of firms who choose to locate in the park (see Lu, Wang and Zhu 2015) and has studied the overall city level treatment effects of building a park (the treatment) in a given year and comparing growth in such treated cities to a set of plausible control cities (Wang 2013). Using data on recent Chinese industrial parks (built between 2004 and 2008), Lu, Wang and Zhu (2015) document that such parks do cause a 12% increase in firm employment, 14.2% increase in output, and also significant increases in capital, capital to labor ratio and labor productivity for firms within parks. Our study differs from this past work along three key margins. First, we focus on the localized spillover effects due to industrial parks based on very fine geographic level data. Second, we are interested in both production and consumption spillovers, so we study how an industrial park triggers the birth of an "edge city" featuring both production and consumption activities. Third, we allow industrial parks to differ along several dimensions – the park's birth cohort and age, the synergy of industries within a park, and how a park "fits" with the local incumbent industries. Estimating the heterogeneous effects of those parks allow us to infer why some parks are successful and some are not in terms of forming a vibrant "edge city" (as a subcenter) far away from the metropolitan area's main center.

Our study builds on research estimating localized productivity spillover effects. Most agglomeration studies focus on the United States (see Rosenthal and Strange 2004, Arzaghi and Henderson 2008, Greenstone, Hornbeck and Moretti 2010). Our new estimates of the co-agglomeration effects builds on the estimates from developed cities generated by Ellison, Glaeser and Kerr (2010). Several studies have examined the role of place based policies on U.S local and regional growth (see Rossi-Hansberg et. al. 2010, Kline and Moretti 2013, Neumark and Kolko 2010). Their general conclusion is that those place-based policies do not lead to net growth, instead, they "reshuffle" economic activities from one place to another place. Such "zero sum" reshuffling suggests that place based policies do not lead to new growth and instead can have perverse consequences by distorting the efficient allocation of resources. However, this argument may not hold in China. Macro research has documented the productivity wedges across Chinese firms (Hsieh and Klenow 2009). If place based policies, such as industrial parks, help to reallocate labor and capital to its highest and best use then local economic growth is likely to result.

Our most novel empirical contribution is to study the consequences of urban sub-center growth on the emerging "consumer city" (see Glaeser, Kolko and Saiz 2001). Urban economists have increasingly been studying cities as centers of consumption. This research has tended to focus on the United States and Europe. A novel feature of our study is to link the rise of the industrial parks (centers of production) to the emergence of new consumer city clusters. While China's cities are known for their high levels of urban air pollution (see Zheng and Kahn 2013), there is nascent rising demand for quality of life in China's cities. Quality of life is both a function of city level local public goods such as climate and air pollution but is also a function of consumer opportunities (Waldfogel 2008). Industrial park workers seek to reduce their commute costs and they are well paid, thus new residential communities and retail opportunities emerge close to the new parks (Lucas and Rossi-Hansberg, 2002). Moretti (2010) estimates such local multipliers using US census data, and finds that for each additional job in manufacturing in a given city, 1.6 jobs are created in the nontradable sector in the same city, and this effect is significantly larger for skilled jobs because they command higher earnings (2.5 for skilled jobs and 1 for unskilled jobs). While past consumer city research has focused on developed nations, one of our paper's contributions is to document the rise of nascent consumer cities in Communist China.

The benefits of place based policies are likely to be localized and heterogeneous (Glaeser and Gottlieb 2008, Kline 2010, Faggio, Silva and Strange, 2015). We present new evidence on the heterogeneous spillover effects of different types of industrial parks. We test several ideas from the economic geography and urban economics literature (Moretti 2004, Rauch 1999). Parks that feature a higher human capital level, a lower level of State Owned Enterprise share and parks that are a better "fit" with the local incumbent industries (in terms of stronger input-output linkages, labor market pooling, and knowledge spillover) have larger local spillovers. Our micro data based approach allows us to test for a rich set of such heterogeneous effects, and to study the key mechanisms generating such spillovers.

Industrial parks are not assigned to geographic locations at random. They are more likely to be built at the suburban fringe where land is cheaper and has not already been developed. We argue that both the invisible hand and the Chinese "managed hand" of the state play a key role as there is a complementarity between highways and fringe industrial parks. Measuring the benefits of active industrial policy remains an open research topic (Beason and Weinstein 1996, Harrison and Rodriguez-Clare 2009). We carefully discuss site selection issues and study how this affects our estimates of productivity and consumption spillovers. We contrast OLS and IV estimates of park spillover effects on local economic activity in which we instrument for the industrial park's location within a city. We argue that OLS and IV estimates of localized spillover effects answer two distinct economic questions. OLS estimates of spillover effects associated with industrial parks built in specific parts of the city provide a historical ex-post evaluation of the effectiveness of past Chinese Communist Party industrial policy investments. In a second set of results, we instrument for where parks are built using topographic, historical land use and population density variables. These instruments proxy for the economic cost of building a park in a given geographic location within a city. By using such cost shifters as instruments, our IV results recover estimates of the productivity effects of park construction in low cost versus high cost areas within a city. As we will discuss below, we control for many observed attributes of different locations such as each location's distance from the city center. Thus, in our IV regressions we will be comparing the productivity dynamics for areas equidistant to the city center that differ with respect to their cost of constructing a new park.

The rest of the paper is organized as follows: In Sections 2 we introduce the institutional background and our conceptual framework; Section 3 describes data and models industrial parks' site selection. Section 4 and 5 present our estimates of the production and consumption spillovers, respectively. We conclude in Section 6.

2. Background and Conceptual Framework

2.1 Industrial Parks and Local Leaders' Investment Problem

Local leaders who seek to be promoted within the Chinese political system have strong incentives to invest in projects that yield local economic growth (Wu et. al. 2013, Zheng et. al. 2014). The goal of establishing industrial parks is to promote FDI, international trade, technological innovation and employment. Industrial parks are at different levels: state-level, provincial-level, and prefecture (or below)-level. Higher-level parks enjoy more favorable policies, such as lower interest rate loans, larger tax, land price and utility price discounts. In this paper we focus on state-level and provincial-level parks.[2] Being a host city of such parks has become a favorite strategy of city mayors to compete for FDI and foster local economic growth.

[2] The reason that we only focus on state-level and provincial level parks is because many of the lower-level industrial parks did not obtain formal approval from the central and provincial governments and violated the relevant laws and regulations. In 2003, the central government did a thorough investigation into industrial parks regarding the violation of land use regulation and other regulations. A large number of those lower-level industrial

In choosing whether to invest in a park, local leaders face a trade-off between a stream of benefits an industrial parks offers and the upfront and opportunity costs of establishing the park. To create a park, a local leader acquires a large parcel of vacant land (or relocates current residents). Industrial parks are highly land intensive. This land must be acquired and this land has a major opportunity cost as developers could build new residential towers or commercial properties. Land closer to the city center is already occupied and is expensive to acquire. Suburban fringe farmland represents a much more attractive location for building parks due to the low land conversion cost there.[3] Once the leader acquires the land, he invests to improve the infrastructure including; paved roads, highways, sewage system, the supply of water, power, gas, cable and phone services and some landscaping. The leader also introduces preferential policies for potential park entrants to attract productive firms, such as tax break and customs duty exemption, discounted land-use fees, and special treatment in securing bank loans (Wang 2013, Adler 2013).

Chinese local leaders are powerful but capital constrained. If they are able to raise enough funds, they can build a park quickly. They raise funds through various channels. On average, bank loans, land sale revenue and on-budget fiscal revenue account for 28%, 25% and 33% of the sources of municipal infrastructure investment in China in 2010, respectively, and the rest of the money comes from other non-bank financial channels.[4] Some local leaders borrow heavily. An industrial park in Liaoning province plans to invest 4.2 billion yuan (US$680 million) in 2015, and almost 100% of this investment comes from bank or trust loans.[5]

Each industrial park has an Administrative Committee (AC), which, on behalf of the city government, takes the responsibility to direct and to administer the park – such as project approval, local taxation, land management, finance, personnel, and public service provision. Once an industrial park is built, the AC leaders will actively solicit FDI and domestic firms with the potential to be highly productive.

---

parks were abolished after this investigation. See detailed documentation in Cartier (2001) and Adler (2013).

3 Such farmland was owned by rural villages. If the city government (the upper government of those rural villages) wants to build a new industrial park on the farmland, it needs to covert the use type of that land parcel from agriculture use to industrial use, and compensate the rural village. In most cases the range of compensation for farmers for land taken is quite low because it is often based on income generated in agriculture use instead of being tied to its opportunity cost (the value of the land if allocated to urban use) (Ding and Song 2005).

4 See: Ryan Rutkowski. Four Myths about local Infrastructure Investment in China. China Economic Watch, the Peterson Institute for International Economics. http://blogs.piie.com/china/?p=3281

5 See: http://www.wantchinatimes.com/news-subclass-cnt.aspx?id=20140816000014&cid=1203

The Park's AC has a plan for the park's industrial composition and the anchor industries (for instance car producers or biotech companies). A recruitment strategy will intend to maximize the industrial synergies both within the park and in a vicinity of the park. Private negotiations take place between the AC and the potential entrants regarding the exact bundle of subsidies each firm will receive if it agrees to enter the park. The AC will offer additional incentives to those firms that they desire the most.

This recruitment process bears a close resemblance to the anchor tenant issue that arises in filling a successful shopping mall (Pashigian and Gould 1998, Gould, Pashigian and Prendergast 2005) and the development of U.S suburban planned towns (Henderson and Mitra 1996). Both the owner of a mall and the Chinese local leaders has a strong incentive to internalize the spillovers taking place within their respective borders. A key difference between mall owners and local leaders is that mall owners only gain profit from activity within their mall. In U.S edge cities, developers only internalize the spillover gains taking place within their land parcel's area. If there are positive spillover effects from a mall to the local community, the mall owner has no claim to those. In contrast, Chinese local leaders have an incentive to internalize these additional outside the park spillover effects, because their goal is to maximize the total output from all the economic activities within their political boundary. In the case of U.S edge cities, the edge city competes with the center city and may make choices that do not maximize the synergies with the incumbent center city (Henderson and Mitra 1996, Orfield 2011). In the Chinese case, the city mayor is in charge of the entire jurisdiction (including the center city and the edge city) and thus internalizes all of the possible synergistic effects. The main purpose of this paper is to estimate those spillovers and explore the heterogeneity in such spillovers with respect to park attributes, and the extent to which this park "fits" with the local incumbent industries.

2.2 Concentrated Purchasing Power Generated by a New Industrial Base

We will estimate the localized production and consumption spillovers that an industrial park generates in its vicinity. Here we provide a simple framework to motivate our empirical approach. The goal of our empirical work is quantify the "chain reaction" created by the introduction of a new industrial park.

Our starting point is that industrial parks attract productive firms, and also improve the productivity of incumbent firms located close and within the park. Such parks will have no aggregate impact on urban growth if their creation simply leads to a spatial reshuffling of economic activity that would have located somewhere in the same city had the park not been created. Wang (2013) presents a cross-city study documenting the positive overall effect of China's industrial parks on urban growth. Such a city/macro study abstracts away from internal workings of a city and that study did not examine how the real estate market or the retail sector is affected by the growth of a major suburban employment center.

Productivity spillovers

Each incumbent manufacturing establishment in a city is endowed with a production function. All else equal, firms located closer to the city center (where the main agglomeration takes place) are more productive. Standard urban economics logic posits that a geographic location's distance to the city center is a sufficient statistic for measuring this agglomeration force. Once an industrial park is built, each location within a city becomes two dimensional as we track the plant's distance to the city center and its distance to the closest industrial park. If a city has multiple parks, other parks will also have some effect on this plant. In the empirical analysis we will construct a variable to measure this global impact of all the parks in the city, but now we abstract from this and only consider the closest park.

An incumbent firm chooses the amount of labor, *L* (priced at a competitive wage *w*); capital, *K* (with input price *p*); and land, *LAND* (priced in the competitive land market with rent *r*), to maximize its corresponding profit Π. *A* is the productivity shifter (TFP), and is used to capture the agglomeration economies this firm enjoys. Output's price is normalized to one.

$$Max_{L,K,T}\ \Pi = f(A, L, K, LAND) - wL - pK - rLAND \qquad (1)$$

Here we allow agglomeration externalities, *A*, to depend on the firm's distance to the central business district (CBD) and the nearest park $D = (distance\_cbd, distance\_park)$, and the park's economic power, $power_{park}$ (here we assume the economic power of the CBD is given). In our empirical work below, we model a park's economic power as a function of its size of economic activity, its average human capital level, the share of SOEs in the park, and the park's co-agglomeration index. This power index is an emergent attribute of a park that depends on how many firms and which firms choose to enter the park, and this is based on those firms' private negotiation with the park's AC. Such firms face a discrete choice decision concerning which city they locate in and which part of a city they choose to locate.[6]

$$A = A(distance\_cbd, distance\_park; power_{park}) \quad (2)$$

Where $\partial A / \partial distance\_cbd < 0; \partial A / \partial distance\_park < 0$; $\partial A / \partial\ power_{park} > 0$

The positive agglomeration economies increase the productivity of nearby incumbent plants. This will lead to higher output, $\partial f / \partial A > 0$, and thus higher profit, which will trigger the entry or relocation of firms who are interested in gaining access to such spillovers to the vicinity of the park. Incumbent firms enjoying increased productivity because of their proximity to the new park should also be less likely to exit. The incumbent firms experiencing productivity growth will expand by occupying more land and hiring more workers. The growth of the park itself and the subsequent entry of firms and firm expansion in the park's vicinity lead to competition for inputs, so labor and land prices will rise.

$$\partial w / \partial A > 0, \quad \partial r / \partial A > 0$$

The increases in both employment and wage in and around the park create the market potential base for the emergence of the consumer city sub-center, as described below.

The Emergence of an "Edge City" : Measuring Consumption Spillovers

Cities are centers of both production and consumption (Glaeser, Kolko and Saiz 2001). Handbury and Weinstein (2014) demonstrate that big cities offer greater variety

[6] We are unable to observe the bargaining game that has taken place between a productive firm and competing mayors. Our analysis should be viewed as a conditional analysis. Given that the AC has filled the park, we can study the industrial composition of the park and its consequences on local production and consumption. We will contrast how parks differ in their productivity spillover effects as a function of their composition but we will not be able to answer questions pertaining to whether the park's composition would be much different if the AC had been slightly more aggressive in competing against other cities' ACs in recruiting firms.

of consumer goods that effectively offer lower consumer prices for specific varieties than what can be found in smaller cities. When there is a density of consumer purchasing power, there is a profit incentive for niche suppliers to co-locate to supply such goods (Waldfogel 2008). The creation of a large industrial park creates a spatially concentrated center of employment and purchasing power. Such new workers are likely to seek out short commutes to work and this creates incentives for real estate developers to bid for land and to develop housing towers in close proximity to the industrial parks. As more people live and work in this new "edge city" and given that these individuals are well paid, the opportunities for retail and restaurants soar and this should trigger entry. These spillover effects of the park both in terms of new housing construction and new retail activity represent an important consumer city agglomeration.

We use the concept of "market potential" (Hanson 2005) to describe the rise in purchasing power around the industrial park, which is the demand base for both the housing and retail markets in this edge city:

$$market\_potential_{edge\ city} = \sum_l employment_l \cdot wage_l \cdot e^{-d_{l,park}} \qquad (3)$$

Where $employment_l$ and $wage_l$ are the employment and average wage in community l, and $d_{l,park}$ is the distance between community l and the park. So our market potential measure is the distance weighted purchasing power of communities in the edge city and the closer communities to the park have larger weights.

We first consider new home sales and home prices. The monocentric model of urban economics predicts that home prices decline with distance from the city center to compensate workers for a longer commute to the city center. Zheng and Kahn (2008) report evidence from Beijing documenting support for this claim. If the new industrial parks generate significant increase in the edge city's market potential, then home prices should rise, controlling for distance to the city center.

$$house\_price = house\_price(distance\_cbd, distance\_park; market\_potential_{edge\ city}) \qquad (4)$$

$$\partial house\_price / \partial\, market\_potential > 0$$

and $\partial house\_price / \partial D < 0,\ D = (distance\_cbd, distance\_park)$.

Home sales will also rise with the market potential in the edge city. We posit that we will observe larger increase in home sales in the communities closer to the park where housing price appreciation is higher:

$$\partial house\_sales / \partial distance\_park < 0$$

Since an industrial park is always located in the city fringe where housing supply is more elastic, we expect that the increase in home sales is relatively smaller than the increase in home prices.

The retail sector will also respond to the growth of a new suburban cluster. In the Bresnahan and Reiss (1990, 1991) model, market demand *retail* is determined by the number of consumers (in our context, *employment* in the edge city), and a representative consumer's demand function, *demand*, which is a function of the income and the retail price (*retail_price*). Translating this relationship using our notation yields:

$$retail_{edge\ city} = retail(market\_potential_{edge\ city}, retail\_price) \quad (5)$$

Their demand specification assumes that if the number of consumers doubles, total market demand will double at any given price; and if consumers become richer, total market demand will also increase. In our case, the retail market in a city is segmented into many local submarkets, denoted by the two distances D (d_{cbd}, d_{park}). As the employment and wage in the edge city around the park increases, the demand in the local retail market in the park's vicinity will also rise. Thus, we should observe more retail store openings in those communities closer to the park

$$\partial \Delta retail / \partial distance\_park < 0$$

Waldfogel (2008) extends this model to allow for preference heterogeneity. If richer consumers have a preference for higher quality restaurants and stores then they have an incentive to co-agglomerate because such niche variety stores face fixed costs to opening and will only open if the demand for their niche is large enough to at least cover the fixed costs. If a park has high human capital workers, more sophisticated richer workers and there is a purchasing power to attract top restaurants and this makes living and working there even more attractive as an endogenous consumer edge city emerges (Diamond 2012).

3. Data

Our study focuses on eight major cities in China: Beijing, Shanghai, Shenzhen, Tianjin, Dalian, Wuhan, Xi'an and Chengdu. These cities include all three first-tier cities in China (Beijing, Shanghai, Shenzhen) and a couple of the top second-tier cities.

3.1 Four Geocoded Data Sets

We construct four key geocoded data sets for these eight cities: (1) 120 state- and provincial-level industrial parks listed in the 2006 "Bulletin List for the Official Boundaries of Chinese Industrial Parks" (See Figure 1),[7] (2) manufacturing plants in the Annual Survey of Industrial Firms (ASIFs) dataset from 1998 to 2007. This survey is conducted by National Bureau of Statistics of China (NBSC); (3) All newly-built residential complexes developed by real estate developers from 2006 to 2013. These data are obtained from each city's local housing authority; (4) New openings of retail establishments during 2006 and 2013 for the categories of restaurants, entertainment facilities and retail shops. These data are obtained from *dianping.com,* which is China's leading local retail information and commentary platform, like *yelp.com* in the US.

*** Insert Figure 1 about here ***

Due to data availability constraints, our housing and retail data cover the time period 2006 to 2013 while the industrial parks data cover the years 1998 to 2007. This means that what we observe is how "people follow jobs", rather than how "jobs follow people", in edge cities.

The Spatial unit of analysis

Given that we seek to quantify localized spillover effects, we must construct consistent geographical units for studying the spatial distribution of industrial plants, housing towers and retail stores. Within a Chinese city, there are three levels of administrative units (from the upper to the lower level): district (or county), *jiedao* ("zone" thereafter) and *juweihui* ("small zone" thereafter). Beijing has 16 districts, 320

[7] This information is supplied by the Ministry of Land and Resources of China (MLRC);

zones and 5274 small zones. In the eight cities, the average sizes of a zone and a small zone are 47.9 and 4.1 square kilometers, respectively.

We know the exact geographic boundaries of industrial parks and zones, but we only have the centroid (rather than its geographic boundary) of a small zone. For all plants in our data set, we know their zone identifiers, and for about 60% of them we know their small zone identifiers. We know the exact street address of all residential complexes and retail establishments. We set up 2 km*2 km grid maps (each cell is 4 square kilometers, the same as the average size of a baby zone) for all eight cities, so that we can count the number of new home sales and new retail openings by grid cell.

Using a major industrial park in Beijing ("Beijing Economic and Technological Development Zone") as an example, Figure 2 shows the above geographic units of analysis we use when doing the geocoding work.

*** Insert Figure 2 here ***

The Industrial Parks Data Set

According to the 2006 "Bulletin List for the Official Boundaries of Chinese Industrial Parks", there are 120 state- and provincial-level industrial parks in these eight cities, accounting for 8.6% in all such parks in China. From the list we know each park's name, location, and the year this park was established. From the websites of industrial parks' ACs we obtain the exact geographic boundary of each park (for those parks that do not have websites or public released information, we contacted the local officials in their ACs to get the boundary information). We then geocode the exact boundaries of all the 120 parks in the GIS maps of the eight cities (Figure 1). Each city has several industrial parks – Beijing has 21 parks, and Dalian has 8 parks. The average park's size is 11.88 square kilometers.[8]

Chinese cities have experienced four waves of establishing industrial parks: 1978 – 1985, 1986 – 1990, 1991 – 1995, and 1996 – 2008 (Wang 2013). The first three waves happened before the start year of our manufacturing plants data set, and the last wave

[8] The largest park is 64, and the smallest one of 0.2 square kilometers. If we measure the distance from the centroid of a park to the corresponding city's CBD, the average park is located 24.9 kilometers away. The most remote park is 95.1 kilometers from the city center.

almost overlaps with our data set. So we define the parks established before1996 as "old parks", and those established in or after 1996 as "new parks".

The Manufacturing Plants Data Set

We obtain plant-level data from the Annual Survey of Industrial Firms (ASIFs) dataset conducted by National Bureau of Statistics of China (NBSC) from 1998 to 2007. All the state-owned enterprises (SOEs) and non-state owned enterprises with annual sales of more than 500 million RMB in the manufacturing sector are surveyed, with detailed information on a plant's identification, operations and performance, and all financial variables. Those firms hire roughly 70% of the industrial employment, generate 90% of the industrial outputs and 98% of the exports (Brandt et al. 2012). An advantage of this ASIFs data set (compared to the economic census data set) is that it enables us to estimate plant-level total factor productivity (TFP). We use the data on outputs and intermediate inputs, deflated by output and input price indices in Brandt et al.(2012), to calculate real capital stock, real value added, and then estimate plant-level TFP. We link plants over time using their information on ID, name, industry code, small zone/zone identifier, etc., and construct an unbalanced panel of 58,834 plants in this ten year period for these eight cities.

To measure the spillovers of these industrial parks in their vicinity, we map plants into industrial parks[9], so that we are able to identify whether a plant is located inside or outside the industrial parks. Figure 3 shows that industrial parks do emerge as subcenters in a city, with much higher employment density (number of jobs per square kilometer) both within parks and in their vicinity, compared to other places in the city. We also compute the average human capital level (years to schooling) by park by year. Figure 3

[9] This involves two processes: geocoding the boundaries of the 120 industrial parks in those eight cities' GIS maps, and geocoding plants in the ASIFs dataset in the GIS maps, so that these two can be merged. We use China's Streetmap GOOGLE online to do the geocoding work. We start with the official street addresses of an industrial park and its land use drawings, and then use ArcGIS to precisely identify the location and geographic boundary of an industrial park. We exploit the address information of each plant in the ASIFs dataset and geocode the exact locations of these plants in GIS maps. We match ASIFs address information with the street maps using village (or *Juweihui*) names and township (or *Jiedao*) names. The ASIFs dataset provides the village-(or *Juweihui*-) level codes for each firm for the period of 2004-2007. To get the village-(or *Juweihui*-) level codes for firms for the period of 1998-2003, we pursued a number of ways. First, we track the firm's village- (or *juweihui*-) level code for firms in the ASIFs data set in the period of 1998-2002 by taking advantage of the information on village (or *Juweihui*) names. Second, because some firms have no information on village (or *Juweihui*) names, we utilize the information on township (or *Jiedao*) names to get their corresponding township- (or *Jiedao*-) level codes. Third, if some firms have neither village (or *Juweihui*) names nor township (or *Jiedao*) names especially in 2003, we turn to Baidu, a local google searching engine, to find their address information by using firms' name information. Then we repeated the first step or the second step to get their village-(or *Juweihui*-) level codes or township- (or *Jiedao*-) level codes.

shows that the human capital level in industrial parks and the zones within 2 km and 5 km from parks are significantly higher than that in the rest of the city.

*** Insert Figure 3 about here ***

We construct two key variables from this plant data set for our empirical analysis. The first variable is the plant-level productivity measure – total factor productivity (TFP). Figure 3 shows that industrial parks and their vicinity do show significantly higher TFP. The average TFP within parks is even higher than that in CBD.[10]

The second variable is the park-specific co-agglomeration index. Ellison and Glaeser (1997) developed a methodology of industry pair EG co-agglomeration index, which measures the extent to which the two industries tend to co-locate in a certain area.[11] We run a simple regression to see how an industrial park's co-agglomeration index evolves over time as the park gets older. The polynomial regression generates the curve presented in Figure D-2 in Appendix D. For the average industrial park, as more firms enter the park, the industry synergy increases over time. This suggests that local governments use their "managed hand" to recruit those firms they want, and this increases the degree of industrial synergy within a park. In the Technical Appendix A we provide the details of how we construct these two variables.

The Housing Data Set

We obtain the price and quantity data for all newly-built residential complexes developed by real estate developers from 2006 to 2013 from the local housing authorities in these eight cities. This data set has the information of the average transaction price and the number of units sold by residential complex by month. The sample size varies from 8,000 to 40,000 complex-month in the eight cities. The

[10] One possibility is that, if incumbent firms anticipate that a park will open up soon. In this case, they may take actions to accumulate inputs and then their TFP could be artificially low in the period just before the park opens. If this is true, we will overestimate parks' effect in generating TFP spillovers. In Figure D-1 in Appendix D we plot the TFP trends of incumbent firms which had existed at least two years before the introduction of an industrial park nearby. We do not observe any TFP drop before the park establishment.

[11] Our park-specific co-agglomeration index is the weighted average of the bilateral EG co-agglomeration indices for the existing industry pairs in the park (using employment in each industry pair as the weight). Intuitively, if those industry pairs that have higher EG co-agglomeration indices have larger employment shares in a park, this park will have a higher park-specific co-agglomeration index, which means that the industries in the park enjoy a higher level of synergy. We calculate the park-specific co-agglomeration index by year using the employment data of three-digit level manufacturing industries in China during 1998-2007.

physical attributes for each complex include the floor area ratio (*FAR*), greening space rate (*GREEN*), and ratio of parking space to the number house units (*PARKING*).

Using data on the longitude and latitude of each residential complex, we geocode all the residential complexes. We calculate each complex's locational attributes such as distance to the city center, distance to the closest industrial park, and which 2 km*2 km grid cell this complex is in. We also calculate for each grid cell, in each month the number of units sold. As shown in Figure 3, there are peaks of both quantities and prices in the housing market in the vicinity of the new industrial parks.[12] These figures highlight that the new parks are creating new city sub-centers.

The Consumption Amenities Data Set

We construct a dataset for local private consumption goods based on dianping.com which is China's leading local retail information and commentary platform, like *yelp.com* in the US. The website covers 12 general categories and nearly 200 sub-categories. The three biggest categories are restaurants, entertainment facilities, and retail shops. For these three major categories, there are more than 990 thousand retail establishments in these eight cities as of 2013. We know the establishment dates of those shops. We geocode them in the GIS maps and calculate each category's density in each grid cell. Figure 3 shows that all three types of consumption amenities have their density peaks in the vicinity of industrial parks.

Table 1 provides definitions and summary statistics of the key variables we use in this paper. Unlike previous localized spillover studies, we have introduced a more comprehensive approach that captures the impact of new parks on the productivity in the manufacturing sector, as well as on the residential sector and the retail sector. Such coverage is crucial because our ASIF dataset contains high-tech companies that produce high-tech goods, such as Intel, Samsung, IBM, Lenovo, Siemens, etc., but it does not contain those high-tech companies that produce Internet services such as Baidu and Alibaba. By explicitly studying the residential and retail sectors, our empirical approach captures these impacts.

[12] There is a limited supply of residential land within industrial parks (most of the land is zoned as industrial land) so many managers and workers buy or rent their apartments in nearby places.

*** Insert Table 1 here **

3.2 Evidence on the Industrial Park Site Selection Decision

In a city, industrial parks are not randomly assigned to locations. In a large city featuring durable capital and farming activity at the fringe, specific geographic areas within the city will differ with respect to land use patterns, infrastructure, and industrial agglomeration. The cheapest and easiest land to create into a park is likely to be at the suburban fringe but such land may be disconnected from the rest of the city and offer few synergistic effects with existing activity. The mayor has a strong incentive to tradeoff these facts when he decides where to place a new park.

Consider an extreme case in which the mayor has perfect foresight concerning what will be the spillover effects from building a new park in each feasible location within the city. In this case, the site selection choice will reflect the anticipated treatment effect. Consider two different models of site locational choice. In the first case, the mayor embraces a complementarity model of economic growth where he places the park in the strongest part of the city. In this case, a naïve econometrician will over-state the true causal impact of a new park because the park has been place in an area with excellent unobservable productive attributes. In a second case, consider a mayor who engages in "spatial compensation" such that he places the park in the area with the weakest fundamentals. He might pursue this strategy to reduce spatial economic inequality. In this second case, the naïve OLS econometrician would under-state the impact of parks because they have been systematically placed in areas with weak fundamentals.

Chinese local leaders are powerful but capital constrained. Land is one key input in production that they can access because of their ability to seize it from farmers located at the suburban fringe. Mayors recognize that industrial parks are land intensive. When deciding on where to locate an industrial park, the land is the major component in the mayor's upfront out-of-pocket cost. To reduce relocation costs and the compensation paid to the original residents on the land parcel where the industrial park will occupy, the local leader tends to choose a low-density place at the city edge, with a large share of farm land. In this way both the relocation cost (paid to original residents) and the compensation cost (paid to villages who own the farm land) will be relatively low. We create our instrumental variables based on this logic. One is the developed land share in 1980 (*DEVELPED_LAND%_1980*), revealed by the remote sensing map of that year. The resolution of this remote sensing map is 1 km*1 km grid cells. From the map we can calculate the share of developed land in each cell, and we aggregate this share to the zone level. The higher this share is, the less likely this zone will be chosen by the local leader for an industrial park due to the high land acquisition and compensation costs. The second instrumental variable is historical population density by zone in 1982 (China's 3rd population census), *POP_DENSITY_1982*. Higher historical population density also means higher compensation cost if this area is converted to an industrial park. Unfortunately this historical population data is unavailable in Shenzhen and Xi'an, so the observations in these two cities will be dropped when this IV is included.

Using the 1980 remote sensing maps, we create our third instrumental variable, *COMMUNIST_LAND%_1980*, which measures the share of land that was designated to big public projects (such as dams, power plants, etc.) owned by China's Communist Party (CCP) and CCP military uses in the old time before China's economic reform in 1980s. There were no land market and other factor markets at that time so the location decisions of such land uses were solely based on CCP's central orders. Therefore this variable tracks the persistence of past CCP land use orders for major public projects. We expect that the higher this "communist land use" ratio, the less likely this zone will be converted to an industrial park later due to the high engineering cost of such conversion.

We use the topographic data to construct our fourth instrumental variable, *FLAT_LAND%*, which measures the share of land with slope smaller than 15 degrees. Building industrial parks on relatively flat land will incur smaller engineering cost, since many parks contain large-scale factory buildings, warehouses, and also high-rise office buildings. The validity of the above four IVs relies on the assumption that those cost measures are only correlated with the probability of building an industrial park at a certain location, but are not correlated with the productivity potential at this location.

We run two versions of probit models to study the correlates between where a park is built as a function of location attributes. In the first version, we are interested in if zone *k* is the home to at least one park in 2006. We estimate Equation (6), where *Z* contains our instrumental variables, and *X* are zone-level controls. The series of *X* are location attributes of zone *k*, such as the distances to the CBD, the railway station, the airport and the closest university, which are all predetermined locations long time before the industrial park was built. We do not include the distance to the highway because a vast highway construction was built in this period so we cannot identify whether the highway affects industrial park location or highways are built because a new park has been built. We include city fixed effects.

$$prob.(whether\ zone\ k\ is\ home\ to\ park(s)) = \alpha_0 + \sum_l \alpha_{1l} \cdot Z_{lk} + \sum_h \alpha_{2h} \cdot X_{hk} + city\ FEs + \varepsilon_{kt} \quad (6)$$

In the second version (Equation (7)), we focus on the 27 industrial parks established in our study period (year 1998-2006). We match each park with all the zones in that city, so the observation here is a zone-park pair. We estimate a conditional logic model to examine whether a newly-established park j matches with zone k. In addition to including the vector of instrumental variables, we are also interested in whether the local leader locates the park in a zone whose original industrial composition has a higher synergy with the proposed industry composition in the park. Such a complementary strategy would maximize the future possible agglomeration economies. To test this, we construct a co-agglomeration index between zone *k*'s original industrial composition (in the initial year) and the park *j*'s industrial composition in the end year, $coagglomeration_{jk}$, and include it in this conditional logic model.

$$\begin{aligned} prob.(whether\ park\ j\ matches\ with\ zone\ k) = \alpha_0 + \sum_l \alpha_{1l} \cdot Z_{lk} + \sum_h \alpha_{2h} \cdot X_{hk} \\ + coagglomeration_{jk} + city\ FEs + \varepsilon_{jkt} \end{aligned} \quad (7)$$

Table 2 presents the estimates of our park selection models. Columns (1) and (2) reports the regression results of the probit model Equation (6). Industrial parks are more likely to locate in the zones further away from CBD where farmland is more available and the opportunity cost is also small. At the same time, industrial parks are more likely to locate close to the railway stations and universities. Such locations have productivity advantages. Zones with larger share of developed land in 1980, larger share of flat land, or higher population density in 1982, are less likely to have industrial parks in 2006. The "communist land use" share in 1980 has the right sign but is statistically insignificant. The joint F test for the *Z* vector (which we will use as instruments below) is statistically significant at 1% level.

In the second version of park site selection model (columns (3) and (4) in Table 2), we introduce our zone-park co-agglomeration index ($COAGGLOMERATION_{jk}$). This variable is not statistically significant. This implies that, at least on average, local leaders do not consider the incumbent firms' synergy with the proposed industrial park when choosing the location for that park. The three variables, *DEVELOPED_LAND%_1980*, *FLAT_LAND%* and log(*POP_DENSITY_1982*) perform well in this conditional logit model. Again, the joint F test for all IVs is statistically significant at the 1% level.

*** Insert Table 2 about here ***

Based on the coefficients of the four instrument variables in column (2) in Table 2, we can construct a "propensity score index" for park location:

$$\textit{propensity score for park location for zone } k = \sum_l \hat{\alpha}_{1l} \cdot Z_{lk} \quad (8)$$

The higher this index, the higher probability (predicted by these four instrumental variables) that zone *k* receives a park. In Figure 4, we map this propensity score for Beijing and we contrast it with the actual park locations within the city. A darker color means that the propensity score is higher. We observe that parks do locate in those places with higher scores. We also plot each zone's average TFP against this propensity score index in Figure 4. There is no clear correlation between these two. The red dots are those zones with parks.

*** Insert Figure 4 about here ***

4. Estimating Productivity Spillovers

4.1 Average TFP Spillovers: OLS and IV Results

We focus on TFP when measuring of local productivity spillovers. We will also examine firm exit probability. We will then study the dynamics of the spatial distribution of local aggregate manufacturing employment activity both within and around industrial parks. This latter outcome measure is useful for judging the increased localized demand for new real estate and consumption opportunities.

Within-Park TFP Premiums

Before estimating the TFP spillovers of industrial parks, we first provide evidence that parks do attract productive firms, and also improve the incumbent firms' TFP. Lu et. al. (2015) present evidence that industrial parks do improve firm employment, output, capital, capital to labor ratio and labor productivity, but they do not directly test the TFP effect due to data constraint.[13] The existence of this within-park TFP premium is the base for our further exploration of TFP spillovers. We estimate Equation (9) for all plants in our data set:

$$\begin{aligned}\log(TFP_{it}) &= \alpha_0 + \alpha_1 \cdot \log(DISTANCE\_CENTER_i) + \alpha_2 \cdot X_i + \alpha_3 \cdot PARK_{ij} + \alpha_4 \cdot AFTER_{jt} \\ &+ \alpha_5 \cdot PARK_{ij} \cdot AFTER_{jt} + district \times year\ trend + industry\text{-}year\ FEs + \varepsilon_{it}\end{aligned} \qquad (9)$$

Where the subscript *i, j, t* refers to plant *i*, industrial park *j* and year *t*. In this equation and all the equations below, $DISTANCE\_CENTER_i$ is the real travel distance (the driving distance based on the road network in that city) from plant *i* to the city center.[14] X_i is a vector containing the plant-specific variables we control for, including whether this plant is a *SOE*, and this plant's distances to the closest railway station, airport and university. $PARK_{ij}$ equals to 1 if plant *i* is located in park *j*. $AFTER_{jt}$ equals to 1 if park *j* exists in year *t*. So this is a DID specification and our main interest lies on the coefficient (α_5) of

[13] Lu et. al. (2015) assume that all of the benefits are within the park and their control group are the nearby firms. Our paper provides the evidence that those nearby firms receive significant spillovers from parks. So their control group is also treated and they underestimate parks' effect.

[14] We use a new travel-distance algorithm written by John Voorheis, a PhD candidate in Economics at the University of Oregon, to construct this real travel distance measure. This STATA code is available on http://pages.uoregon.edu/jlv/code.html.

the interaction term $PARK_{ij} \cdot AFTER_{jt}$. We recognize that this coefficient could be positive due to both selection and treatment effects.

We define the spatial unit of analysis as follows; when we study industrial plant spillovers we define a plant is within an industrial park if its small zone's centroid is located within the park's boundary (if this plant only has zone identifier then we use that zone's centroid) and the park exists in that year.

Table 3 presents five estimates of equation (9). Column (1) is estimated using OLS and includes all plants with either zone or small zone identifiers. We observe a significantly negative gradient for TFP with respect to the distance to the city center. SOE plants have a TFP discount of about 24.4%. All else equal, those plants in industrial parks have a 22.7% TFP premium, which is a sum of both selection and treatment effects. We then only keep those plants with small zone identifiers, and 62% of the plants are left. In column (3) we include plant fixed effects to explicitly model the firm selection effect. The TFP premium shrinks from 22.9% (column (2)) to 14.6% for this subsample of incumbent plants. By including plant fixed effects, we are identifying the park effect based on the subsample of incumbent firms that were located in the industrial park's area before and after the park opened. The fact that the park's fixed effect shrinks highlights that the ACs are recruiting productive firms.

In columns (4) and (5) we report IV estimates where we instrument for whether a zone has a park using our historical land use, historical population density, and land slope as IVs. The joint F tests for IVs in the first stage are all statistically significant. We can observe that the TFP premium for within-park plants rises from 22.9% (based on the OLS results) to 27%-35%. The fact that our productivity estimates rise when we instrument for the spatial placement of the park (and the instrument is based on identifying low cost areas within the city to place the park) suggests that the mayors are either unaware of the differential benefits of placing a park in one area or another within the city or they are intentionally seeking to place the parks in less productive parts of the city. This is consistent with our finding in Table 2 that the zone-park co-agglomeration index does not matter in the park site selection model.

*** Insert Table 3 about here ***

Estimating TFP Spillovers

We estimate a similar equation as Equation (9) but here we only keep those plants located outside of the industrial parks. We measure a plant's proximity to a park by its distance to the closest park (log(*DISTANCE_PARK*)). In this equation and all the equations below, *DISTANCE_PARK* is measured as the real travel distance based on the road network in that city (see footnote #14 for how we construct this variable). This is a dynamic variable – it will shrink if a new park is located near the plant.

We recognize that in a major city that local productivity could also be affected by the main city center (CBD) and other industrial parks in the same city. The CBD and other parks increase the city's overall size and this would increase the city's economic growth. We use the distance to CBD (*DISTANCE_CENTER*) to control for the global impact of the city's main center. We construct a variable (*GLOBAL_PARK_IMPACT*) to measure this global impact of all the parks in the city. We borrow the idea in the market potential studies to construct this variable:

$$GLOBAL\_PARK\_IMPACT_{it} = \sum_{j \neq j0} w_{ij} \cdot \log(EMP_{jt}) = \sum_{j \neq j0} \left[1 - \left(\frac{d_{ij}}{d_{\max}} \right)^2 \right]^2 \cdot \log(EMP_{jt}) \qquad (10)$$

In each year *t*, we identify all the existing parks in the city, and the closest park *j0*. For all the parks except of park *j0*, we use the inverse distance (in quadratic weighting function) from this plant *i* to those parks as weights to compute the weighted sum of those parks' employment in year *t*. If a city has more parks and the plant is relatively closer to those parks, we expect this plant will receive more spillovers and thus have a higher TFP. We will control for this "global impact" of parks, and focus on the gradient of the spillovers from the closest park. We estimate Equation (11).

$$\begin{aligned}\log(TFP_{it}) = {} & \alpha_0 + \alpha_1 \cdot \log(DISTANCE\_CENTER_i) + \alpha_2 \cdot X_i + \alpha_3 \cdot \log(DISTANCE\_PARK_{ij}) \\ & + \alpha_4 \cdot GLOBAL\_PARK\_IMPACT + district \times year\ trend + industry \times year\ FEs + \varepsilon_{it}\end{aligned}$$

(11)

The coefficient α_3 measures the spatial decay rate of the park's TFP spillovers, i.e., $\partial A / \partial DISTANCE\_PARK < 0$ in our conceptual framework. Panel A in Table 4 reports the regression results. This spatial decay rate is -0.056 for all plants (column (1)) and is -0.044 for the plants with small zone identifiers (column (2)), and both are OLS regressions. Note that this "local" TFP gradient generated by the closest park is 1.1 to 1.6 times the "global" TFP gradient generated by the CBD. In columns (3) we include plant fixed effects, and this decay rate shrinks a little bit to -0.041. So the firm selection issue (a park's AC selects more productive firms into the park) does not induce a big overestimation of this local TFP gradient. Besides the "local" impact of the closest park, the "global" impact of all other parks in the city is also statistically significant. One standard deviation increase in *GLOBAL_PARK_IMPACT* can induce a 0.12 standard deviation increase in TFP.

Columns (4) and (5) report IV regressions. This decay rate becomes steeper (-0.124 and -0.19) compared to that in column (2). Now this local TFP gradient generated by the closest park is 4 to 6 times the gradient from the CBD. Consistent with the results presented in Table 2, this OLS-IV comparison also indicates that, instead of putting an industrial park in the most productive parts of the city, city mayors tend to place them in less productive places in the city. We note that this finding is based on controlling for the distance to the city center so we are comparing geographic areas in a concentric circle around the city center.

*** Insert Table 4 about here ***

If industrial parks generate positive productivity spillovers in their vicinity, we should observe that, all else equal, firms closer to the park will have a higher probability to survive. We test this by estimating the following probit model:

$$\begin{aligned} prob.(whether\ plant\ i\ exits) = \alpha_0 + \alpha_1 \cdot AGE_i + \alpha_2 \cdot \log(DISTANCE\_CENTER_i) + \alpha_3 \cdot X_i \\ + \alpha_4 \cdot (DISTANCE\_PARK_{ij}) + City\ FEs + \varepsilon_{it} \end{aligned}$$

(12)

The observation in this regression is plant *i* in our firm data set. AGE_i is the age of plant *i* since its appearance in our firm data set. If it exits from our data set during our study period, the dependent variable equals to 1, otherwise it equals to 0. There may be several reasons for a plant's exit: closure, relocation, or the size of its total output shrinks to far below the threshold of the ASIFs survey (500 million RMB).[15] Since we do not know the reasons for why individual firms exit the sample, we assume that all sample exits represents a firm's death. Table A2 in the Appendix C reports the regression results of this probit model. Column (1) is for all plants. Older plants have a higher probability to be closed. We find that plants closer to industrial parks have a lower probability to exit. This is another piece of evidence documenting the industrial parks' production spillovers. Interestingly, those plants closer to the city center, universities and railway station have a higher probability to be closed. One possibility is that those places are undergoing urban redevelopment and industrial land uses are replaced by commercial and residential land uses, so many manufacturing firms are relocated to other places.

4.2 Heterogeneity in TFP spillovers

We now test for industrial park TFP spillover heterogeneity by allowing the TFP spatial gradient to vary along four dimensions. The first dimension is based on the age of the park. We have two cohorts – old parks established before 1996 and new ones established in or after 1996. If the Chinese leaders built the best parks first and if diminishing returns have taken place for subsequent parks then we will observe that the new parks have weaker spillovers than the old ones. We find evidence of diminishing returns, the spatial gradient of the old parks' TPF spillovers is 2.3 times that of the new parks (-0.171 vs. -0.075). Within each cohort, our panel data set of manufacturing plants' TFP allows us to examine as a park gets more mature, does its spillover effect get larger. Panel B in Table 4 reports such heterogeneous spillovers. Within the old park cohort, as a park matures over time, its TPF spillovers get stronger. For the new parks, those with an age between 11-15 years also have much stronger spillovers than younger ones. The regression has the same specification as column (3) in Panel A (OLS regression with district-time trend and industry-year fixed effects).

The second heterogeneity depends on the park's own economic power. We construct five "park power" indicators ($power_{park,m}$, m=1,2,3,4,5): (1) park's distance to the CBD.

---

[15] Plants whose sales slip to a small extent below this threshold are not automatically removed from this sample since the 5 million RMB is not a 'hard' rule (see details in Brandt et al. (2012))

Closer distance means stronger linkage with the city's mean economic center so more powerful; (2) park's size. Larger parks are more powerful; (3) the share of SOE employment in the park. SOEs are less productive so smaller share of SOEs means more powerful parks; (4) the average human capital level (years to schooling) in the park. Higher human capital means more powerful; (5) the co-agglomeration index for the industries residing in the park, higher synergy between those industries will lead to a more powerful park. Our hypothesis is that more powerful parks (measured by these five indices) generate larger spillovers.

The third heterogeneity dimension measures the extent to which this park "fits" with the local incumbent industries. Following Glaeser and Kerr (2009), Ellison et al. (2010) and Jofre-Monseny et al. (2011), We construct four park-vicinity "synergy indices" ($synergy_{park\text{-}vicinity,n}$, n=1,2,3,4) to measure the input-output linkages, the size of labor market pooling and the knowledge spillover possibility between within-park and outside-park firms. In Appendix B, we report the details of how we construct these four indices. Our hypothesis here is that if an incumbent firm in a park's vicinity has higher synergy indices with the park, it will enjoy a larger spillover effect from the park.

The fourth heterogeneity depends on plant attributes. We construct two variables to measure a plant's size and age. Size is measured in a plant's employment. We group all plants into three age categories based on their establishment year: before 1978 (before the economic reform), between 1979 and 1998 (earlier stage of the economic reform) and after 1998 (later stage of the reform).

We estimate Equation (13) to (15) to explore the heterogeneity in the spatial decay rate of TFP spillovers with respect to the second to the fourth sets of heterogeneity measures.

$$
\begin{aligned}
\log(TFP_{it}) = {} & \alpha_0 + \alpha_1 \cdot \log(DISTANCE\_CENTER_i) + \alpha_2 \cdot X_i + \alpha_3 \cdot \log(DISTANCE\_PARK_{ij}) \\
& + \alpha_4 \cdot GLOBAL\_PARK\_IMPACT + \sum_{m=1}^{4} (\alpha_{5,m} \cdot power_{park,m} \cdot \log(DISTANCE\_PARK_{ij})) \\
& + district \times year\ trend + industry \times year\ FEs + \varepsilon_{it}
\end{aligned}
$$

(13)

$$\log(TFP_{it}) = \alpha_0 + \alpha_1 \cdot \log(DISTANCE\_CENTER_i) + \alpha_2 \cdot X_i + \alpha_3 \cdot \log(DISTANCE\_PARK_{ij})$$
$$+ \alpha_4 \cdot GLOBAL\_PARK\_IMPACT + \sum_{n=1}^{5}(\alpha_{5,n} \cdot synergy_{park-vicinity,n} \cdot \log(DISTANCE\_PARK_{ij}))$$
$$+ district \times year\ trend + industry \times year\ FEs + \varepsilon_{it}$$

(14)

$$\log(TFP_{it}) = \alpha_0 + \alpha_1 \cdot \log(DISTANCE\_CENTER_i) + \alpha_2 \cdot X_i + \alpha_3 \cdot \log(DISTANCE\_PARK_{ij})$$
$$+ \alpha_4 \cdot GLOBAL\_PARK\_IMPACT + \alpha_5 \cdot PLANT\_SIZE_{it} \cdot \log(DISTANCE\_PARK_{ij})$$
$$+ \sum_{l=2}^{3}(\alpha_{6l} \cdot AGE\_GROUP_{it,l} \cdot \log(DISTANCE\_PARK_{ij}))$$
$$+ district \times year\ trend + industry \times year\ FEs + \varepsilon_{it}$$

(15)

Table 5 reports the heterogeneity TFP spillovers results. In the first three columns we interact the four park-vicinity synergy indices with log(*DISTANCE_PARK*). In column (1) with all plants, higher input linkage and output linkage, as well as stronger skill spillover between within-park and outside-park plants will lead to larger park TFP spillovers. We only keep those plants with small zone identifier in column (2) and then further include plant fixed effects in column (3), and the output linkage index is statistically significant. All the four synergy indices are jointly significant in the three columns, indicating that those parks that are better "fit" into their vicinity in terms of industry synergies can generate higher TFP spillovers. In columns (4)-(6) we interact the five park power indices with log(*DISTANCE_PARK*). Distance to CBD does not influence this gradient significantly, while the other four power indices have some impacts and have the right signs. These five interaction terms are also jointly significant, showing that more powerful parks lead to higher spillover effects. In columns (7)-(9) we interact plant size (employment) and age group dummies with log(*DISTANCE_PARK*). Smaller and younger plants enjoy higher TFP spillovers from nearby parks.

*** Insert Table 5 about here ***

5. Estimating Consumption Spillovers

5.1 The Increase in Consumer Market Potential in Edge Cities

The productivity increase both inside the park and in its vicinity will attract new entrants, and the incumbent firms will also expand their employment. A growing workforce that seeks a short commute to work will seek out housing and retail opportunities near the new industrial park. All together the market potential will increase and thus trigger new housing construction and retail activities in those edge cities.

To measure these effects, we estimate equations (16) and (17) where the dependent variable is the manufacturing employment density *EMP*. The unit of analysis here is zone *k*. In estimating these regressions, in one specification we use the data for all zones and in a second we focus on the subset of zones outside parks. For the latter, we control for the global impact of other parks in the city, and focus on the impact of the closest park. See Equation (16) and (17).

$$EMP_{kt} = \gamma_0 + \gamma_1 \cdot \log(DISTANCE\_CENTER_k) + \gamma_2 \cdot PARK_{kj} + \gamma_3 \cdot AFTER_{jt} + \gamma_4 \cdot PARK_{kj} \cdot AFTER_{jt} + District\ FEs + Year\ FEs + \varepsilon_{kt} \quad (16)$$

$$EMP_{kt} = \gamma_0 + \gamma_1 \cdot \log(DISTANCE\_CENTER_k) + \gamma_2 \cdot \log(DISTANCE\_PARK_{kjt}) + \gamma_3 \cdot GLOBAL\_PARK\_IMPACT_{kt} + District\ FEs + Year\ FEs + \varepsilon_{kt} \quad (17)$$

Table 6 studies how the presence of an industrial park leads to local manufacturing employment growth. All columns are zero inflated poisson (ZIP) regressions. Columns (1) – (3) include all zones, and columns (4) to (6) only includes those zones located outside of parks. Column (1) shows that, all else equal, a zone with an industrial park features 78% higher manufacturing employment density. In the IV regression (column (2), joint F test for IVs in the first stage are statistically significant), this employment density gap is even higher—about twice of the counterpart's density. We do find industrial parks significantly influence the nearby employment density, and the spatial decay rate is about -52% (both in logs, column (4)), and it is 1.3 times the decay rate from the CBD. The size of this decay rate is smaller in the IV regression than in the OLS regressions (columns (4) and (5)). As column (3) and (6) show, an industrial park's power in generating local employment growth is larger for larger parks, and for those parks with higher human capital, a larger co-agglomeration index and lower share of SOE firms. The global impact of other parks on the local employment is insignificant.

*** Insert Table 6 about here ***

5.2 Home Sales and Pricing

In many major Chinese cities, the center city is the most vibrant part of the metropolitan area. Similar to major European cities, jobs, government, and amenities are clustered downtown (Brueckner, Thisse and Zenou 1999). Given the attraction of the city center, real estate prices are quite high close to the downtown and decline with distance from the downtown (Zheng and Kahn 2008).

Chinese mayors seek to encourage economic growth in other parts of the metropolitan area. As we documented in Table 2, a typical industrial park is located at the edge of the urban area whose previous use as agriculture land. The emergence of this new edge city and the rising purchase power of consumers will generate new demand for housing and local consumption goods, and thus trigger housing development and the opening of new retail stores nearby, as shown in Figure 3. The main land use type within industrial parks is industrial land, so managers and workers in those plants will buy or rent residential units outside but near the park. That is why in Figure 3 we see the closest zones to parks have the highest densities of housing transactions in the surrounding area, even higher than that in parks.

To empirically test whether there is a booming housing market around the park, we construct two dependent variables – the number of new home sales in the 2 km*2 km grid cell *g* by quarter *($HOUSE\_SALES_{gt}$)*, and the transaction price in a residential complex *l* by quarter ($HOUSE\_PRICE_{lt}$). We use each complex's exact street address to calculate its shortest real travel distance to the closest park's boundary. We expect that, controlling for the real travel distance to the main center (CBD), those grid cells closer to a park will see more new home sales, and we should also observe that homes closer to the park sell for a price premium. We control for the global impact of other parks in the city. See Equation (18) and (19).

$$\begin{aligned} HOUSE\_SALES_{gt} = \gamma_0 + \gamma_1 \cdot \log(DISTANCE\_CENTER_g) + \gamma_2 \cdot \log(DISTANCE\_PARK_{gj}) \\ + \gamma_3 \cdot GLOBAL\_PARK\_IMPACT_{gt} + district\ FEs + year\ FEs + \varepsilon_{gt} \end{aligned} \qquad (18)$$

$$\begin{aligned} \log(HOUSE\_PRICE_{lt}) = \gamma_0 + \gamma_1 \cdot \log(DISTANCE\_CENTER_l) + \gamma_2 \cdot \log(DISTANCE\_PARK_{lj}) + \boldsymbol{\varphi} \cdot X_l \\ + \gamma_3 \cdot GLOBAL\_PARK\_IMPACT_{lt} + district\ FEs + year\ FEs + \varepsilon_{lt} \end{aligned}$$

(19)

We are only able to trace back the housing transaction data to 2006, and at that time all industrial parks had been built, therefore what we observe are the consequences of industrial parks on nearby housing market, rather than the reverse effect from housing market to parks.

The estimates of the housing quantity equation (Equation (18)) are reported in columns (1) – (3) in Table 7, which are ZIP regressions. In column (1), after controlling for distance to the CBD, we observe a negative gradient with respect to the distance to the closest park (as a subcenter). The size of this spatial decay rate is quite large, about -0.221 (log-log). Note that the price gradient with respect to the distance to the CBD is insignificant here. In the ZIP-IV regression (column (2)), this decay rate becomes smaller (-0.199) , and it is statistically significant at 1% level, indicating there may be some unobserved location attributes that are positively related to both residential development activities and the location choice of an industrial park. The global impact of other parks on local housing sales is also statistically significant but the economic size is small.

Column (3) tells us that residential units close to the parks further from the CBD, and the parks with higher human capital and higher co-agglomeration index are more likely to be sold out. The five heterogeneity measures of a park's power are jointly significant.

Columns (4) – (6) report hedonic housing price regressions (Equation (18)). We control for a residential complex's physical attributes such as its construction density (floor-to-area ratio), parking space share and green space ratio (X_l). We find a significantly negative gradient of housing price with respect to the distance to the nearby park. This housing price decay rate is about -0.101 (log-log) in column (4) (OLS), and its absolute value also shrinks in the IV regression in column (5) (-0.095). The magnitude of this local gradient is about 55% of the global price gradient (-0.170) with respect to the distance to the CBD. The global impact of other parks on local housing prices is also statistically significant and has a big economic size– a standard deviation increase in *GLOBAL_PARK_IMPACT* can induce a 0.55 standard deviation increase in local housing sales. Larger parks, parks with higher co-agglomeration index, higher human capital, less SOE employment are more attractive to home buyers. The five heterogeneity measures of a park's power are also jointly significant.

*** Insert Table 7 about here***

Retail Activities

We explore how industrial parks impact local retail opportunities, using data on three major types of local consumption goods, restaurants, retail shops and entertainment facilities. We seek to test whether industrial parks have triggered a "snowball" gentrification effect and created a vibrant consumption neighborhoods in edge cities. We run Equation (20) for the new openings of each of these three local consumption goods, and the unit of analysis here is grid cell/year.

$$\Delta RETAIL_{gt} = \delta_0 + \delta_1 \cdot \log(DISTANCE\_CENTER_g) + \delta_2 \cdot \log(DISTANCE\_PARK_{gj}) + \delta_3 \cdot GLOBAL\_PARK\_IMPACT_{gt} + district\ FEs + year\ FEs + \varepsilon_{gt}$$

(20)

The regression results of Equation (20) are presented in Table 8. Columns (1), (4) and (7) are ZIP regressions; (2), (5), (8) are IV regressions (joint F tests for IVs are statistically significant in the first stage); and (3), (6), (9) are ZIP regressions with interaction terms. Those grid cells closer to industry parks do see more openings of fancy restaurants, shops and entertainment places. And again, the spatial decay rates of these three consumption amenities shrink in IV regressions. The local gradient generated by the closest park is about a quarter of the global gradient generated by the main center for all the three amenity types. The global impact of other parks on the quantity of consumption amenities is statistically significant but the economic sizes are small. We again find that parks further away from the CBD and parks with larger size have larger gentrification effects. The five heterogeneity measures of a park's economic power are also jointly significant for these three local private goods.

*** Insert Table 8 about here ***

Recall that the coefficients of *DISTANCE_PARK* in TFP equations become larger in IV regressions compared to those in OLS regressions. Here for housing and retail equations, we find such coefficients all shrink from OLS to IV regressions in TFP equations. This indicates that, city mayors locate parks in those places with larger consumption potential, but smaller production potential. One possible explanation for city mayors' intension behind this park location decision is that, city mayors can receive higher land sale revenue by selling the residential and commercial land parcels near industrial parks if the parks' consumption spillovers are large. Housing and retail are non-tradable goods. Instead, city mayors have to set very low price for industrial land parcels in parks and also provide tax break in the initial years to attract mobile firms. Chinese city mayors have a term constraint (four years for one term. A mayor always stay in position for one to two terms). From the mayor's perspective, the revenue from residential and commercial land sales can be realized in a lump-sum form in the short-run, while tax revenue from productive firms can only come in annually in the long-term, so they have the incentive to place parks in those places with higher consumption potential.

6. Conclusion

At the broadest level, our paper provides strong support for Marshallian theories of localized production and consumption co-agglomeration in a leading developing country. Using the opening of over 120 industrial parks across eight major Chinese cities as a natural experiment, we quantify the spatial decay patterns in productivity, firm survival, real estate construction, real estate pricing, and retail store openings for economic activity close to these new suburban centers of productivity. Consistent with Marshall's core hypotheses, we find that proximity to the parks reduce the costs of moving goods, people, and ideas.

Unlike other agglomeration studies, we have explicitly integrated our analysis of the production and consumption side of the economy. The new park creates a spatially concentrated increase in local market potential as well paid workers seek nearby housing and retail opportunities. We have provided new evidence on this local multiplier effect in a developing nation (Moretti 2010). This local multiplier effect is larger for those industrial parks that attract more human capital and more private firms.

We have explicitly modeled the tradeoffs that the mayor faces in siting a new park in a specific area within a city. A forward looking mayor has an incentive to recognize that each area in a city may offer different benefits and impose different costs if a park is built there. When we use cost shifter proxies as instrumental variables, we estimate a larger TFP treatment effect from building a park. But, we estimate smaller IV estimates in our housing and retail regressions. This collective set of facts suggests that mayors locate parks in those places featuring greater consumption potential but lower producer efficiency potential.

One explanation for these facts is due to the Chinese political system. Chinese mayors face term limits so that they serve for at most eight years. From the mayor's perspective, the revenue from residential and commercial land sales can be realized in a lump-sum form in the short-run, while tax revenue from productive firms can only come in annually in the long-term, so they have the incentive to place parks in those places with higher consumption potential.

References

Adler, Solomon. (2013), *The Chinese Economy*, Routledge.

Arzaghi, Mohammad, and J. Vernon Henderson. (2008) "Networking off madison avenue", *The Review of Economic Studies*, 75 (4), 1011-1038.

Beason, Richard, and David E. Weinstein. (1996), "Growth, economies of scale, and targeting in Japan (1955-1990)", *The Review of Economics and Statistics*, 78, 286-295.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. (2012), "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing", *Journal of Development Economics*, 97 (2), 339-351.

Bresnahan, Timothy F., and Peter C. Reiss. (1990), "Entry in monopoly market", *The Review of Economic Studies*, 57 (4), 531-553.

Bresnahan, Timothy F., and Peter C. Reiss. (1991), "Entry and competition in concentrated markets", *Journal of Political Economy*, 99, 977-1009.

Brueckner, Jan K., Jacques-Francois Thisse, and Yves Zenou. (1999), "Why is central Paris rich and downtown Detroit poor?: An amenity-based theory", *European Economic Review*, 43 (1): 91-107.

Cartier, Carolyn. (2001) "Zone Fever, the Arable Land Debate, and Real Estate Speculation: China's evolving land use regime and its geographical contradictions", *Journal of Contemporary China*, 10 (28), 445-469.

Combes, Pierre - Philippe, et al. (2012), "The productivity advantages of large cities: Distinguishing agglomeration from firm selection", *Econometrica*, 80 (6), 2543-2594.

Diamond, Rebecca. (2012), "The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000." Job Market Paper, Harvard University

Ding, Chengri, and Yan Song, eds. (2005), *Emerging land and housing markets in China*. Lincoln Inst of Land Policy.

Ellison, Glenn, and Edward L. Glaeser. (1997), "Geographic concentration in US manufacturing industries: a dartboard approach", *Journal of Political Economy*, 105(5), 889-927.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. (2010), "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review*, 100, 1195-1213.

Faggio, Giulia, Olmo Silva, and William C. Strange. "Heterogeneous agglomeration.", *Review of Economics and Statistics, forthcoming.*

Garreau, Joel. (1991), *Edge cities: Life on the new frontier*, *NY: Doubleday*.

Glaeser, Edward L., and Joshua D. Gottlieb. (2008), "The economics of place-making policies", No. w14373. National Bureau of Economic Research.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. (2001), "Consumer city", *Journal of economic geography*, 1 (1), 27-50.

Glaeser, Edward L., and William R. Kerr. (2009), "Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?", *Journal of Economics & Management Strategy*, 18 (3), 623-663.

Gould, Eric D., B. Peter Pashigian, and Canice J. Prendergast. (2005), "Contracts, externalities, and incentives in shopping malls", *Review of Economics and Statistics*, 87 (3), 411-422.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. (2010), "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings", *Journal of Political Economy*, 118 (3), 536-598.

Handbury, Jessie, and David E. Weinstein. (2014), "Goods prices and availability in cities", *The Review of Economic Studies*, rdu033.

Hanson, Gordon H. (2005), "Market potential, increasing returns and geographic concentration", *Journal of international economics*, 67 (1), 1-24.

Harrison, Ann, and Andrés Rodríguez-Clare. (2009), "Trade, foreign investment, and industrial policy for developing countries", No. w15261. National Bureau of Economic Research.

Henderson, Vernon, and Arindam Mitra. (1996), "The new urban landscape: Developers and edge cities", Regional Science and Urban Economics, 26 (6), 613-643.

Hsieh, Chang-Tai and Peter Klenow, (2009), "Misallocation and Manufacturing TFP in China and India". *Quarterly Journal of Economics*, CXXIV 4, 1403-1458.

Jofre-Monseny, Jordi, Raquel Marín-López, and Elisabet Viladecans-Marsal. (2011), "The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms", *Journal of Urban Economics*, 70 (2), 61-74.

Kline, Patrick. (2010), "Place based policies, heterogeneity, and agglomeration", *The American Economic Review*, 100, 383-387.

Kline, Patrick M., and Enrico Moretti., (2013) "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee valley authority", No. w19293. National Bureau of Economic Research.

Lu, Yi, Wang Jin, Zhu Lianming (2015), Do Place-Based Policies Work? Micro-Level Evidence from China's Economic Zones Program.

Lucas, Jr, Robert E., and Esteban Rossi-Hansberg. (2002) "On the internal structure of cities", *Econometrica*, 70(4), 1445-1476.

Moretti, Enrico. (2004), "Workers' education, spillovers, and productivity: evidence from plant-level production functions", *American Economic Review*, 94, 656-690.

Moretti, Enrico. (2010), "Local Multipliers", *American Economic Review: Papers and Proceedings*, 100: 1-7.

Neumark, David, and Jed Kolko. (2010), "Do enterprise zones create jobs? Evidence from California's enterprise zone program", *Journal of Urban Economics*, 68 (1), 1-19.

Orfield, Myron. (2011), *American metropolitics: The new suburban reality*, Brookings Institution Press.

Pashigian, B. Peter, and Eric D. Gould. (1998), "Internalizing Externalities: the Pricing of Space in Shopping Malls 1", *The Journal of Law and Economics*, 41 (1), 115-142.

Rauch, James E. (1999), "Networks versus markets in international trade", *Journal of international Economics*, 48 (1), 7-35.

Rosenthal, Stuart S., and William C. Strange. (2004), "Evidence on the nature and sources of agglomeration economies", *Handbook of regional and urban economics*, Vol. 4, 2119-2171.

Rossi-Hansberg, Esteban & Pierre-Daniel Sarte & Raymond Owens. (2010), "Housing Externalities", *Journal of Political Economy*, 118 (3), 485-535.

Waldfogel, Joel. (2008), "The median voter and the median consumer: Local private goods and population composition", *Journal of Urban Economics*, 63 (2), 567-582.

Wang, Jin. (2013), "The economic impact of special economic zones: Evidence from Chinese municipalities", *Journal of Development Economics*, 101, 133-147.

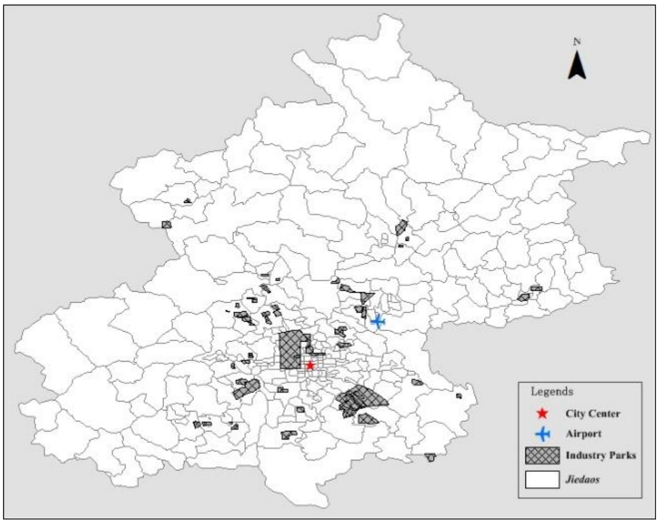
Wu, Jing, et al. (2013), "Incentives and outcomes: China's environmental policy", No. w18754. National Bureau of Economic Research.

Zheng, Siqi, et al. (2014), "Incentives for China's urban mayors to mitigate pollution externalities: The role of the central government and public environmentalism", *Regional Science and Urban Economics*, 47, 61-71.

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

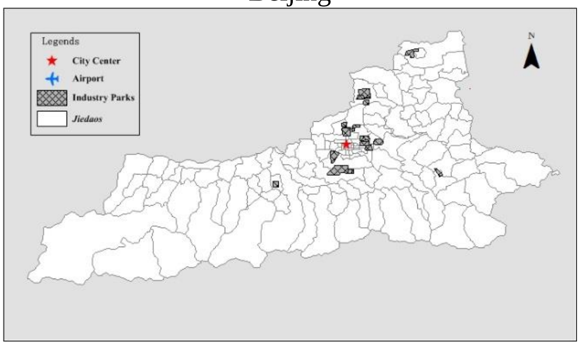
Zheng Siqi, and Matthew E. Kahn. (2013). “Understanding China's Urban Pollution Dynamics”, *Journal of Economic Literature*, 51(3), 731-772.

Figure 1 Within-city locations and geographic boundaries of industry parks

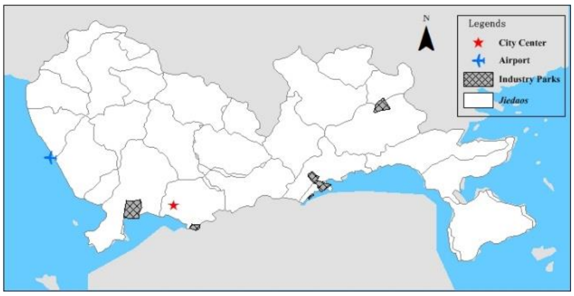


Beijing

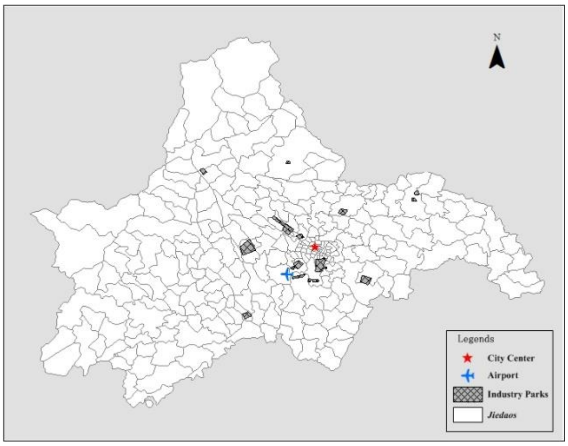
Shanghai



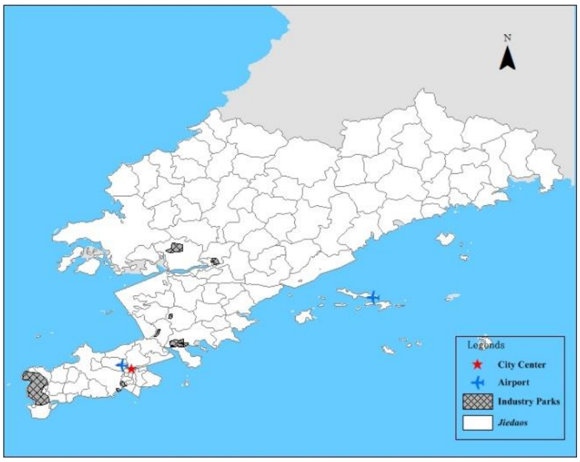
Xi’an



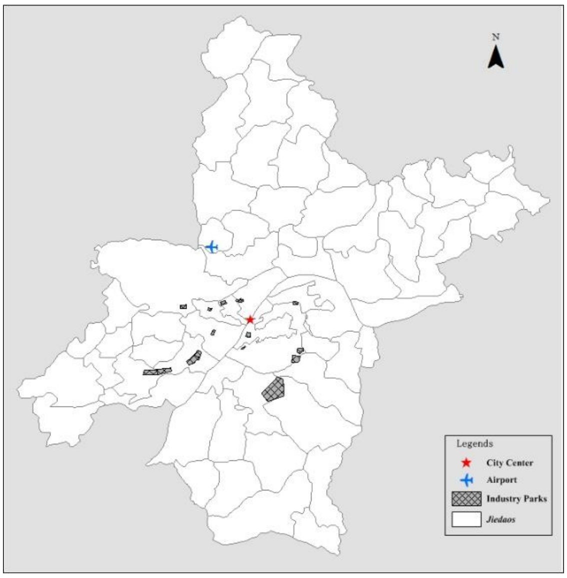
Shenzhen



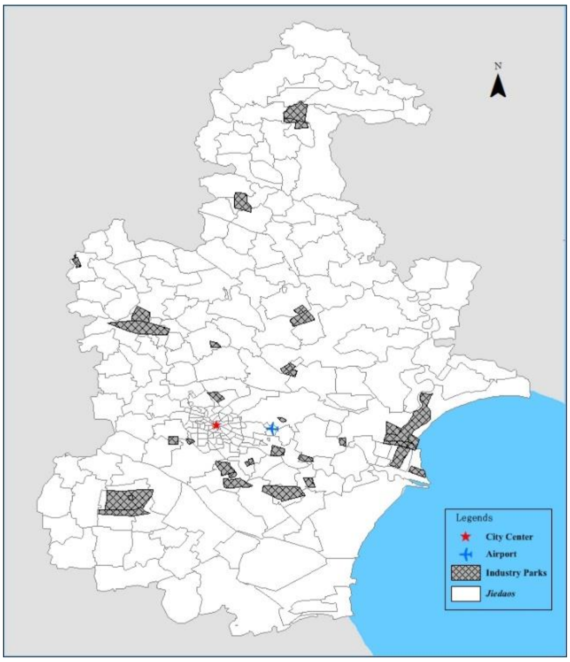
Chengdu



Dalian



Wuhan



Tianjin

Figure 2 Geographic Unit of Analysis

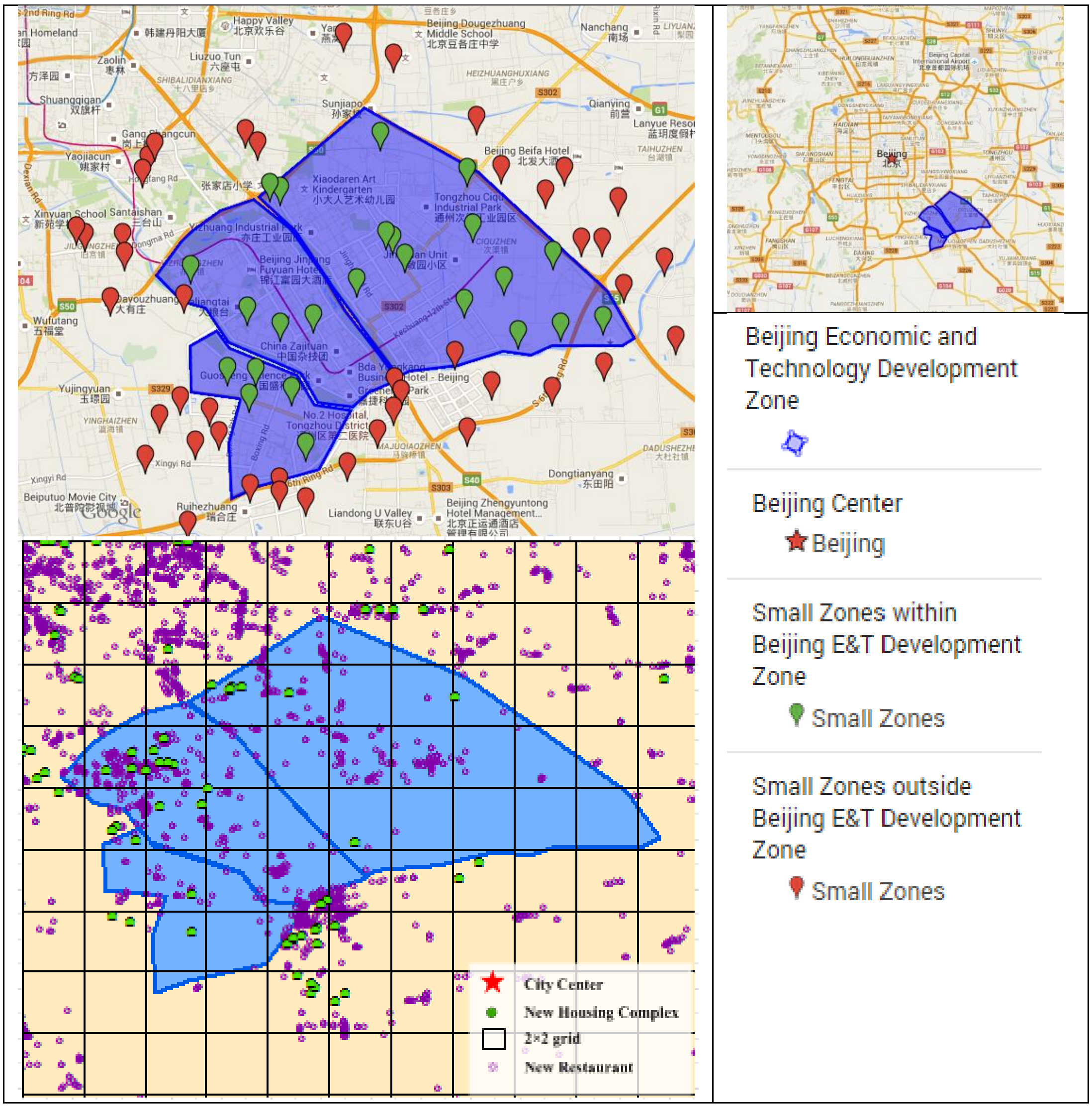
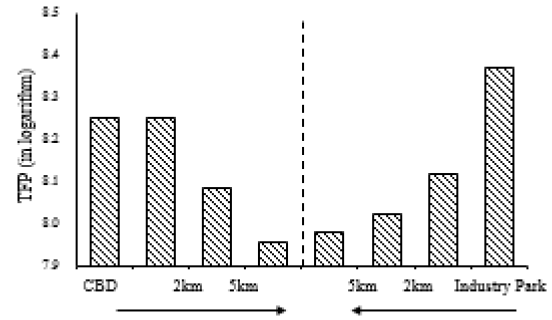
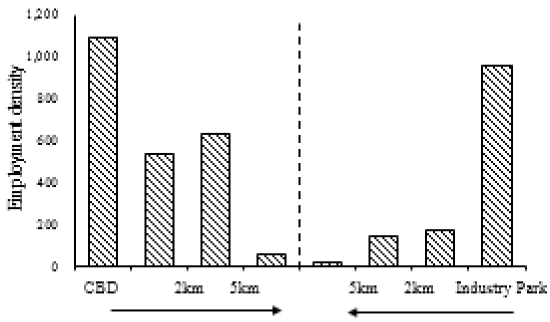


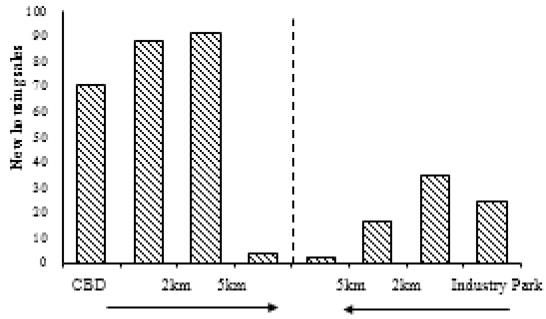
Figure 3 The Spatial Distributions of Key Economic Activity Indicators in Eight Cities



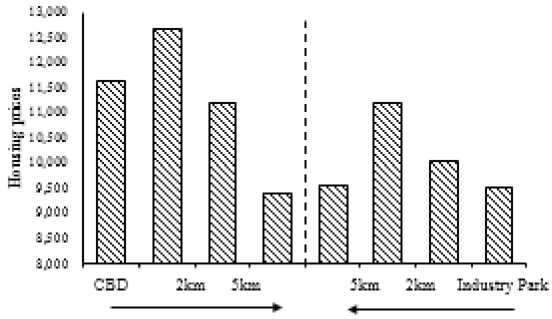
TFP (in logarithm)



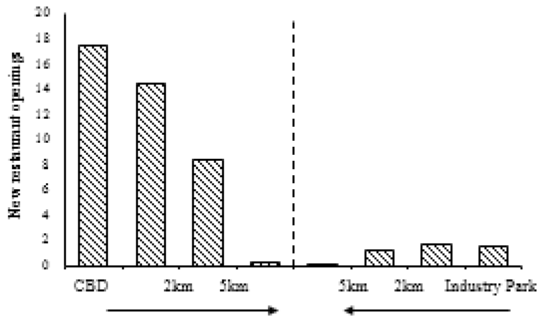
Employment density (workers/km^2)



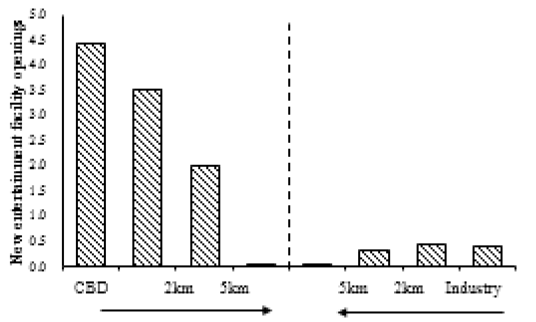
New housing sales (number of units/km^2)



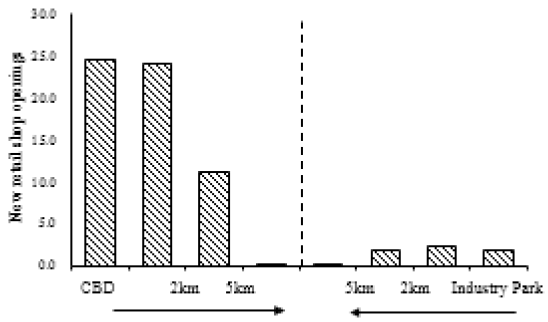
Housing prices (RMB/m^2)



New restaurant openings (number of openings/km^2)

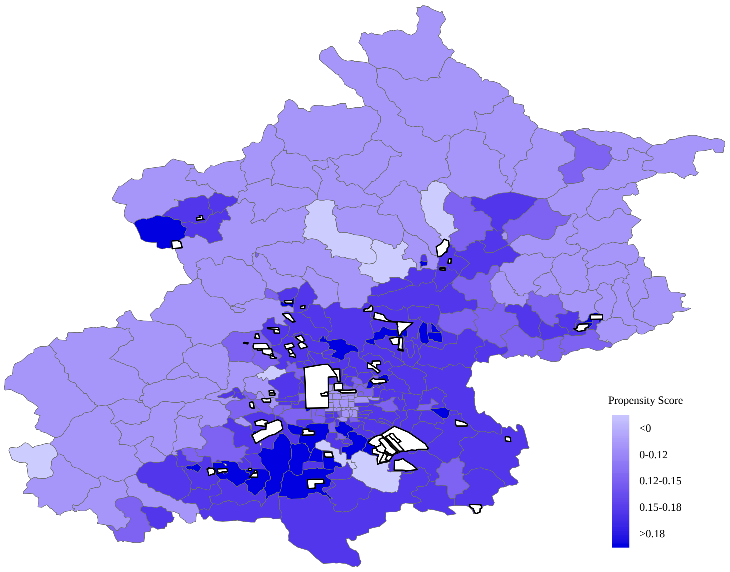


New entertainment facility openings (number of openings/km^2)

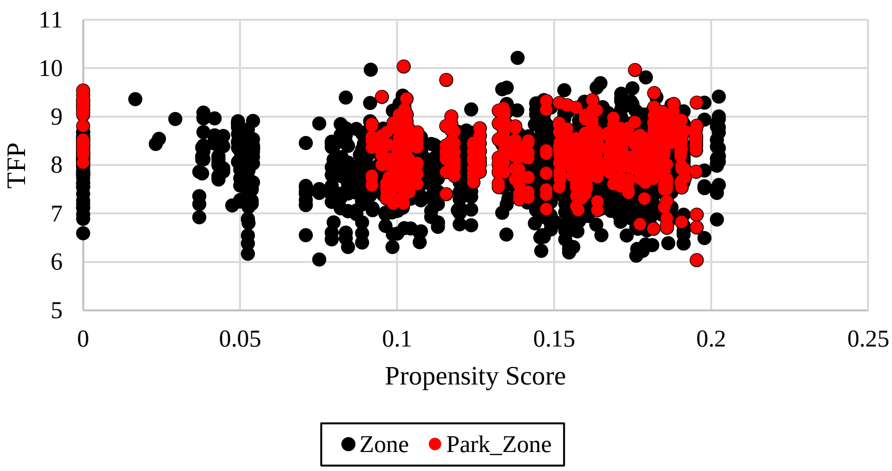


New retail shop openings (number of openings/km^2)

Figure 4. A Propensity Score Index for Predicting a Beijing Park's Location



(1) Propensity score for park location and the real park locations in Beijing



(2) Propensity score for park location and parks' average TFP in Beijing

Table 1 Variable Definitions and Summary Statistics

Variable	Definition	Obs.	Mean	Std. Dev.	Min	Max
Manufacturing data by plant by year						
TFP_{it}	Total factor productivity of plant *i* in year *t* (in logarithm)	179588	8.26	1.18	-2.90	15.57
$DISTANCE\_CENTER_i$	Real travel distance (based on the road network) of plant *i* to the city center (km)	179588	21.20	18.19	0.36	229.620
$PARK_{ij} \cdot AFTER_{jt}$	=1 if plant *i* is in an existing industrial park *j* in year *t*	179588	0.33	0.47	0	1
$DISTANCE\_PARK_{ijt}$	Real travel distance (based on the road network) of plant *i* to the closest industrial park *j* in year *t* (km)	179588	5.68	8.49	0	124.61
SOE_i	=1 if plant *i* is an SOE (State Owned Enterprise)	179588	0.31	0.46	0	1
$PLANT\_SIZE_{it}$	Total employment of plant *i* in year *t* (in logarithm)	179588	4.73	1.15	2.08	11.37
AGE_{it}	Age of plant: *AGE*=1 for the plants established before 1978; *AGE*=2 for those established between 1978 and 1998; *AGE*=3 for those established after 1998	179588	2.35	0.61	1	3
$GLOBAL\_PARK\_IMPACT_{it}$	The global impact of all parks (except the closest one) on plant *i* in year *t*. See Equation (9).	179588	117.82	72.54	0	254.78
$DISTANCE\_RAILWAY_{it}$	Distance to the closest railway station in year *t* (km)	179588	5.96	5.68	0.01	66.64
$DISTANCE\_AIRPORT_{it}$	Distance to the closest airport in year *t* (km)	179588	23.10	13.87	1.21	102.40
$DISTANCE\_UNIVERSITY_i$	Distance to the closest university (km)	179588	14.47	16.94	0.16	190.51
Industrial Park attributes by park by year						
$DISTANCE\_CENTER_{jt}$	Real travel distance (based on the road network) of park *j* to the city center (km)	942	26.67	16.65	4.99	87.29
$PARK\_SIZE_j$	The planned area of park *j* (km^2)	942	13.48	14.59	0.20	64.08
SOE_SHARE_{jt}	Share of SOEs in park *j* in year *t*	942	0.30	0.15	0	1
$HUMAN\_CAPITAL_{jt}$	Share of workers with education attainment of college and above in park *j* in year *t*	942	0.18	0.06	0.08	0.39
$COAGGLOMERATION_{jt}$	Co-agglomeration index of park *j* in year *t* (see Appendix A for how we construct this variable)	942	0.14	0.09	-0.02	0.55
$DISTANCE\_RAILWAY_{jt}$	Distance to the nearest railway in year *t* (km)	942	6.46	4.85	0.40	28.04
$DISTANCE\_AIRPORT_{jt}$	Distance to the nearest airport in year *t* (km)	942	27.72	17.32	2.83	79.38
$DISTANCE\_UNIVERSITY_j$	Distance to the nearest university (km)	942	17.50	16.16	0.46	82.14
Park-vicinity "synergy indices"	(see Appendix B for how we construct these variable)					
$INPUT\_LINKAGE_{rjt}$	Input linkages between park *j* and industry *r* in year t	34800	0.02	0.05	0	0.77
$OUTPUT\_LINKAGE_{rjt}$	Output linkages between park *j* and industry *r* in year t	34800	0.03	0.06	0	0.84
$LABOR\_POOLING_{rjt}$	The size of labor market pooling between park *j* and industry *r* in year t	34800	0.02	0.02	0	0.30
$SKILL\_SPILLOVER_{rjt}$	The knowledge spillover possibility between park *j* and industry *r* in year t	34800	0.02	0.02	0	0.31
IVs for park locational choice by zone						
$DEVELOPED\_LAND\%\_1980_k$	Developed land share in 1980	1689	29.17	32.84	0	100

$COMMUNIST\_LAND\%\_1980_k$	The share of land that was designated to big public projects (such as dams, power plants, etc.) and military uses	1689	1.11	4.51	0.00	84.66
$FLAT\_LAND\%_k$	The share of land with slope smaller than 15 degrees	1689	0.89	0.18	0.10	1.00
$POP\_DENSITY\_1982_k$	Historical population density by zone in 1982	1461	5508.84	13208.12	0	166185.2
Employment density by zone by year						
EMP_{kt}	Manufacturing employment density in zone k in year t (/km^2)	15999	235.86	744.77	0	19075
$GLOBAL\_PARK\_IMPACT_{kt}$	The global impact of all parks (except the closest one) on zone k in year t. See Equation (9).	15999	67.52	52.96	0	253.05
Housing construction and retail activities by 2km×2km grid by year						
$HOUSE\_SALES_{gt}$	Number of new housing sales in grid g in year t	174280	23.29	197.39	0	10826
$RESTAURANT_{gt}$	Number of new restaurant openings in grid g in year t	174280	1.63	15.93	0	1312
$ENTERTAINMENT_{gt}$	Number of new entertainment establishment openings in grid g in year t	174280	0.43	4.27	0	309
$SHOP_{gt}$	Number of new retail shop openings in grid g in year t	174280	2.33	24.81	0	2184
$GLOBAL\_PARK\_IMPACT_g$	The global impact of all parks (except the closest one) on cell g as of year 2007. See Equation (9).	174280	76.26	52.83	1.55	278.71
Residential complex data by complex by year						
$HOUSE\_PRICE_{lt}$	Average housing sale price in complex l in year t (yuan RMB/m^2)	117132	8335.98	5383.29	1000	48003
$DISTANCE\_CENTER_l$	Distance to the city center (km)	117132	13.54	12.43	0.05	152.02
$DISTANCE\_PARK_l$	Distance to the closest industrial park (km)	117132	6.06	5.66	0.55	161.97
$GLOBAL\_PARK\_IMPACT_l$	The global impact of all parks (except the closest one) on complex l as of year 2007.	117132	83.24	63.11	0.46	271.75
FAR_l	The floor area ratio	117132	2.93	2.11	0.06	27.45
$GREEN_l$	Greening space ratio (%)	117132	36.71	9.26	0	95
$PARKING_l$	Parking space share	117132	0.76	0.44	0	6.89

Table 2 The Within City Determinants of Industrial Park Locational Choice

Dependent variable: A Dummy Indicating whether park *j* locates in zone *k*

PANEL A

	Dependent variable =1 If zone *k* is the home to at least one park in 2006		Dependent variable =1 If park *j* established in 1998-2006 located in zone *k*	
	(1)	(2)	(3)	(4)
log(*distance to the closest existing park$_{kt}$*)			-0.228	0.213
			(0.199)	(0.243)
log(*DISTANCE_CENTER$_k$*)	0.031**	0.046***	0.124	0.506
	(0.013)	(0.015)	(0.282)	(0.313)
log(*DISTANCE_RAILWAY$_{kt}$*)	-0.024***	-0.022***	-0.335***	-0.414***
	(0.007)	(0.007)	(0.096)	(0.099)
log(*DISTANCE_AIRPORT$_{kt}$*)	-0.004	-0.001	-0.501**	-0.518**
	(0.0122)	(0.0141)	(0.219)	(0.256)
log(*DISTANCE_UNIVERSITY$_k$*)	-0.059***	-0.058***	0.185	-0.353
	(0.012)	(0.013)	(0.271)	(0.295)
COAGGLOMERATION$_{jk}$			-2.630	-12.060
Between zone k (initial year) and park j (end year)			(8.710)	(14.230)
DEVELOPED_LAND%_1980$_k$	**-1.16e-3*** **	**-1.07e-3** **	**-0.015** **	**-0.021** **
	(4.33e-4)	**(4.42e-4)**	**(0.007)**	**(0.008)**
COMMUNIST_LAND%_1980$_k$	**-0.001**	**-0.001**	**-0.117**	**-0.093**
	(0.002)	**(0.002)**	**(0.095)**	**(0.088)**
FLAT_LAND%$_k$	**0.337*** **	**0.382*** **	**2.823* **	**3.216**
(share of land with slope smaller than 15°)	**(0.064)**	**(0.079)**	**(1.706)**	**(2.449)**
log(*POP_DENSITY_1982$_k$*)		**-0.470*** **		**-4.620* **
		(0.165)		**(2.767)**
City fixed effects	Yes	Yes	Yes	Yes
Joint F-test of IVs	31.91***	36.57***	7.24*	9.89**
	(0.000)	(0.000)	(0.064)	(0.042)
Observations	1689	1461	5234	4769
chi2	170.6	169.3	30.18	30.49

PANEL B: Propensity score index for park location (calculated using the coefficients of the 4 IVs in column (2), see Equation (8))

	Mean	**Std. Dev.**	**Min**	**Max**
Propensity score for park location	0.15	0.04	-0.12	0.21

Note: Panel A reports results from fitting versions of equations (6) and (7). Column (1) and (2) employ probit model to examine whether zone *k* is the home to at least one industrial park by the end of our study period (year 2006), so it is a cross-sectional regression with zones as observations. Column (3) and (4) employ conditional logistic model to examine whether a newly-established park *j* (we only focus on the 27 industrial parks established in our study period, year 1998-2006) matches with zone *k*. Here the observations are zone-park pairs. *COAGGLOMERATION$_{jk}$* is the co-agglomeration index calculated between zone k's original industrial composition one year before park j was established and park j's industrial composition in year 2007. *DEVELOPED_LAND%_1980*, *COMMUNIST_LAND%_1980*, *FLAT_LAND%* and log(*POP_DENSITY_1982*) are exogenous variables for parks' location choice, see the text for how we construct them. Since historical population data is unavailable in Shenzhen and Xi'an, the observations in these two cities are dropped in column (2) and (4). Marginal effects are reported for probit models (column (1) and (2)). Robust standard errors are reported in parentheses, which are clustered by urban district. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Panel B reports the statistics for propensity score index for park location, which is calculated using the coefficients of the 4 IVs in column (2), see Equation (8).

Table 3 The TFP Premium for Plants Located Within Industrial Parks

Dependent variable: TFP_{it} (in logarithm)

	Plants with zone or small zone identifiers	Plants with small zone identifiers			
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	3 IVs	4 IVs
log($DISTANCE\_CENTER_i$)	-0.054***	-0.065***		-0.056***	-0.074***
	(0.019)	(0.023)		(0.010)	(0.011)
$PARK_{ij}$	0.047	-0.008		-0.129	-0.169
	(0.037)	(0.034)		(0.132)	(0.139)
$AFTER_{jt}$	0.005	-0.006	-0.0189	-0.015	-0.051
	(0.025)	(0.025)	(0.0229)	(0.044)	(0.049)
$\boldsymbol{PARK_{ij}\cdot AFTER_{jt}}$	0.227***	0.229***	0.146***	0.269**	0.352**
	(0.043)	(0.037)	(0.0386)	(0.128)	(0.139)
SOE_i	-0.244***	-0.212***		-0.213***	-0.225***
	(0.015)	(0.015)		(0.008)	(0.008)
log($DISTANCE\_RAILWAY_{it}$)	-0.022**	-0.033***	0.0218	-0.032***	-0.034***
	(0.010)	(0.008)	(0.0156)	(0.004)	(0.004)
log($DISTANCE\_AIRPORT_{it}$)	0.011	-0.015	0.0841**	-0.017*	-1.06e-4
	(0.024)	(0.017)	(0.0399)	(0.009)	(0.010)
log($DISTANCE\_UNIVERSITY_i$)	-0.053***	-0.041***		-0.049***	-0.042***
	(0.018)	(0.016)		(0.008)	(0.008)
Constant	7.832***	10.03***	-317.3	10.04***	10.09***
	(0.351)	(0.083)	(199.9)	(0.467)	(0.464)
District-Time trend	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Industry-year fixed effects	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Plant fixed effects	No	No	**Yes**	No	No
Joint F-test of IVs in first stage (p-value)					
$PARK_{ij}$ regression				5.59***	4.88***
				(0.000)	(0.000)
$PARK_{ij}\cdot AFTER_{jt}$ regression				10.70***	7.62***
				(0.000)	(0.000)
Observations	179588	112014	112014	112014	98237
R^2	0.196	0.193	0.712	0.188	0.185

Note: This table reports results from fitting versions of equation (9). All plants both within- and outside-parks are included. It is a DID specification that we include whether plant *i* is located in park *j* ($PARK_{ij}$), whether park *j* exists in year *t* ($AFTER_{jt}$), and the interaction term between these two dummies ($PARK_{ij}\cdot AFTER_{jt}$). Therefore the default category is the TFPs of the out-side plants before the closest park was established. For a plant outside parks, we match it to its closest park and use that park's opening time to assign value to $AFTER_{jt}$. In column (1) we include all the plants (with either zone or sub-zone identifier), in column (2)-(5) we only include the plants with sub-zone identifier so the sample size shrinks by about 38%. District-trend fixed effects and industry-year fixed effects are included in all the regressions. Plant fixed effects are included in column (3). The four exogenous variables in Table 2 (*DEVELOPED_LAND%_1980*, *COMMUNIST_LAND%_1980*, *FLAT_LAND* and log(*POP_DENSITY_1982*)) are used as IVs for $PARK_{ij}$ (and their interactions with $AFTER_{jt}$ as IVs for $PARK_{ij}\cdot AFTER_{jt}$) in our IV regressions. In column (5) the observations in Shenzhen and Xi'an are dropped as the historical population data is unavailable in these two cities. The joint F-tests in column (4) and (5) show that the IVs are jointly significant in the first-stage IV regressions. Robust standard errors are reported in parentheses, which are clustered by zone in column (1) and by small zone in column (2)-(5). * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Table 4 TFP Spillovers in a Vicinity of an Industrial Park

Dependent variable: TFP_{it} (in logarithm)

PANEL A: Average Effects for All Parks

	Plants with zone or small zone identifier	Plants with small zone identifier			
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	3 IVs	4 IVs
log($DISTANCE\_CENTER_i$)	-0.035*	-0.039*		-0.028**	-0.029**
	(0.018)	(0.020)		(0.012)	(0.013)
log($DISTANCE\_PARK_{ijt}$)	-0.056***	-0.044***	-0.041**	-0.124***	-0.190***
	(0.013)	(0.013)	(0.020)	(0.030)	(0.034)
$GLOBAL\_PARK\_IMPACT_{it}$	1.64e-3***	1.84e-3***	1.96e-3***	1.41e-3***	5.84e-4
	(5.25e-4)	(4.61e-3)	(4.82e-4)	(3.21e-4)	(3.69e-4)
SOE_i	-0.250***	-0.232***		-0.232***	-0.248***
	(0.017)	(0.017)		(0.009)	(0.009)
log($DISTANCE\_RAILWAY_{it}$)	-0.021**	-0.030***	-0.026	-0.028***	-0.029***
	(0.008)	(0.008)	(0.019)	(0.005)	(0.005)
log($DISTANCE\_AIRPORT_{it}$)	-0.006	-0.036*	0.051	-0.030***	-0.011
	(0.021)	(0.021)	(0.041)	(0.011)	(0.012)
log($DISTANCE\_UNIVERSITY_i$)	-0.013	0.007		0.016	0.023**
	(0.017)	(0.018)		(0.011)	(0.009)
Constant	7.631***	9.917***	94.74	10.07***	8.315***
	(0.153)	(0.095)	(97.08)	(0.471)	(0.256)
District-time trend	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Industry-year fixed effects	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Plant fixed effects	No	No	**Yes**	No	No
Joint F-test of IVs in first stage				33.48***	24.57***
(p-value)				(0.000)	(0.000)
Observations	123109	79157	79157	79157	66951
R^2	0.187	0.185	0.695	0.185	0.181

PANEL B: Heterogeneous Gradients of TFP Spillovers for Old and New Parks with Different Ages

(The same specification as column (3) in Panel A)

	Average	Park age			
		0-5 years	6-10 years	11-15 years	>15 years
Old parks (built before 1996)	-0.171	-0.106	-0.136	-0.177	-0.150
	(0.096)	(0.114)	(0.098)	(0.094)	(0.095)
New parks (built in or after 1996)	-0.075	-0.073	-0.030	-0.145	-
	(0.024)	(0.024)	(0.026)	(0.054)	

Note: Panel A in this table reports results from fitting versions of equation (11). Only plants outside parks are included. $DISTANCE\_PARK_{ijt}$ is calculated by plant as plant i's distance to the closest industrial park. In column (1) we include all the plants (with either zone or small zone identifier), in column (2)-(5) we only include the plants with sub-zone identifier so the sample size shrinks by about 36%. District-trend fixed effects and industry-year fixed effects are included in all the regressions. Plant fixed effects are included in column (3). The variable GLOBAL_PARK_IMPACT measures the global impact of all the parks (except the closest one, $j0$) on a plant in the city. It is defined as:

$$GLOBAL\_PARK\_IMPACT_{it} = \sum_{j \neq j0} w_{ij} \cdot \log(EMP_{jt}) = \sum_{j \neq j0} \left[1 - \left(\frac{d_{ij}}{d_{\max}}\right)^2\right]^2 \cdot \log(EMP_{jt})$$

, where we use the inverse distance (in quadratic weighting function) from this plant i to those parks as weights to compute the weighted sum of those parks' employment in year t.

We construct the IVs for $DISTANCE\_PARK_{ijt}$ (in logarithm) in the following way: for each IV in Table 2 (*DEVELOPED_LAND%_1980*, *COMMUNIST_LAND%_1980*, *FLAT_LAND* and log(*POP_DENSITY_1982*), we calculate the weighted average of this IV's values in all zones except of the zone where this outside-park firm is located, using the inverse distance (in quadratic weighting function) from this firm's zone to all the other zones as weights: $IV_j(Distance) = \sum_{n \neq j} w_{jn} \cdot IV$, where $w_{jn} = \left[1 - \left(\frac{d_{jn}}{d_{\max}}\right)^2\right]^2$. We use these four weighted average variables as the IVs for this outside-firm's distance to the closest park. In column (5) the observations in Shenzhen and Xi'an are dropped as the historical population data is unavailable in these two cities. The joint F-tests in column (4) and (5) show that the IVs are jointly significant in the first-stage IV regressions. Robust standard errors are reported in parentheses, which are clustered by zone in column (1) and by small zone in column (2)-(5).

Panel B reports the heterogeneous spatial gradient of the TFP spillovers in a park's vicinity. Old parks are those built before 1996, and new parks are those built in or after 1996. For each park cohort, we also look at the heterogeneity in the spatial gradient by park's age (0-5 years, 6-10 years, 11-15 years, older than 15 years). Robust standard errors are reported in parentheses.

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

Table 5 Measuring the Determinants of TFP Spillover Heterogeneity

Dependent variable: *TFP*$_{it}$ (in logarithm). OLS Regressions

	Heterogeneity w.r.t. spillover synergies			Heterogeneity w.r.t. park attributes			Heterogeneity w.r.t. plant attributes		
	Plants with zone or small zone identifier	Plants with small zone identifier		Plants with zone or small zone identifier	Plants with small zone identifier		Plants with zone or small zone identifier	Plants with small zone identifier	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(*DISTANCE_CENTER*$_{i}$)	-0.036**	-0.039*		-0.030*	-0.042**		-0.046***	-0.056***	
	(0.018)	(0.020)		(0.018)	(0.021)		(0.017)	(0.020)	
log(*DISTANCE_PARK*$_{ijt}$)	-0.024*	-0.014	-4.56e-4	0.020	-0.009	-0.182	-0.592***	-0.567***	0.038
	(0.014)	(0.014)	(0.026)	(0.052)	(0.046)	(0.139)	(0.027)	(0.028)	(0.109)
INPUT_LINKAGE$_{ijt}$*log(*DISTANCE_PARK*$_{ijt}$)	-0.121**	-0.023***	-9.67e-4						
	(0.055)	(0.007)	(0.009)						
OUTPUT_LINKAGE$_{ijt}$ *log(*DISTANCE_PARK*$_{ijt}$)	-0.153**	-0.203***	-0.238***						
	(0.078)	(0.075)	(0.076)						
LABOR_POOLING$_{ijt}$ *log(*DISTANCE_PARK*$_{ijt}$)	-0.123	-0.376	0.386						
	(0.258)	(0.290)	(0.267)						
SKILL_SPILLOVER$_{ijt}$ *log(*DISTANCE_PARK*$_{ijt}$)	-0.478*	-0.062	-0.453*						
	(0.266)	(0.307)	(0.272)						
log(*DISTANCE_CENTER*$_{j}$)*log(*DISTANCE_PARK*$_{ijt}$)				-5.85e-3	7.64e-3	0.021			
				(0.008)	(0.008)	(0.045)			
PARK_SIZE$_{j}$*log(*DISTANCE_PARK*$_{ijt}$)				-5.20e-4*	-3.34e-4	-7.88e-4*			
				(3.10e-4)	(2.89e-4)	(4.69e-4)			
SOE_SHARE$_{jt}$*log(*DISTANCE_PARK*$_{ijt}$)				0.352***	0.271***	0.129*			
				(0.078)	(0.076)	(0.073)			
HUMAN_CAPITAL$_{jt}$*log(*DISTANCE_PARK*$_{ijt}$)				-0.844***	-0.765***	-0.053**			
				(0.280)	(0.232)	(0.027)			
COAGGLOMERATION$_{jt}$*log(*DISTANCE_PARK*$_{ijt}$)				-0.060	-0.019	-0.335			
				(0.044)	(0.050)	(0.308)			
PLANT_SIZE$_{it}$*log(*DISTANCE_PARK*$_{ijt}$)							0.054***	0.052***	0.013***
							(0.003)	(0.003)	(0.004)
AGE$_{it}$ **(1978-1998)***log(*DISTANCE_PARK*$_{ijt}$)							0.283***	0.277***	-0.122
							(0.016)	(0.017)	(0.109)
AGE$_{it}$ **(1998-)***log(*DISTANCE_PARK*$_{ijt}$)							0.329***	0.324***	-0.120
							(0.0163)	(0.017)	(0.110)

$GLOBAL\_PARK\_IMPACT_{it}$	1.52e-3***	1.71e-3***	1.88e-3***	1.45e-3***	1.75e-3***	1.98e-3***	1.76e-3***	1.88e-3***	1.91e-3***
	(5.23e-4)	(4.61e-4)	(5.03e-4)	(5.29e-4)	(4.69e-4)	(5.06e-4)	(5.22e-4)	(4.55e-4)	(5.03e-4)
SOE_i	-0.251***	-0.232***		-0.251***	-0.233***		-0.219***	-0.207***	
	(0.017)	(0.017)		(0.017)	(0.017)		(0.016)	(0.016)	
log($DISTANCE\_RAILWAY_{it}$)	-0.021**	-0.031***	-0.037*	-0.019**	-0.030***	-0.041*	-0.024***	-0.033***	-0.041*
	(0.008)	(0.008)	(0.022)	(0.008)	(0.008)	(0.022)	(0.008)	(0.008)	(0.022)
log($DISTANCE\_AIRPORT_{it}$)	-0.006	-0.036*	0.058	-0.003	-0.037*	0.059	-0.005	-0.033*	0.058
	(0.021)	(0.021)	(0.049)	(0.021)	(0.021)	(0.049)	(0.020)	(0.020)	(0.049)
log($DISTANCE\_UNIVERSITY_i$)	-0.013	0.007		-0.012	0.008		-0.017	0.002	
	(0.017)	(0.018)		(0.016)	(0.018)		(0.017)	(0.018)	
Constant	7.657***	9.928***	40.07	7.290***	9.887***	32.93	7.974***	9.841***	39.81
	(0.155)	(0.096)	(182.7)	(0.225)	(0.106)	(182.8)	(0.168)	(0.093)	(182.8)
District-Time trend	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Industry-year fixed effects	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Plant fixed effects	No	No	**Yes**	No	No	**Yes**	No	No	**Yes**
Joint F-test of interaction variables	18.39***	24.21***	9.49***	5.63***	3.77***	2.90**	181.89***	166.50***	3.33**
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.013)	(0.000)	(0.000)	(0.019)
Observations	123109	79157	79157	123109	79157	79157	123109	79157	79157
R^2	0.189	0.187	0.704	0.188	0.186	0.704	0.207	0.203	0.704

Note: This table reports OLS regression results from fitting versions of equation (13)-(15) by including interaction terms between spillover synergy indices with log($DISTANCE\_PARK_{ijt}$) in column (1) - (3), interaction terms between park attributes with log($DISTANCE\_PARK_{ijt}$) in column (4) - (6), and interaction terms between plant attributes with log($DISTANCE\_PARK_{ijt}$) in column (7) - (9). The joint-F tests show that the interaction terms are jointly significant in all columns. See Appendix B for details about how we construct the spillover synergy indices. Robust standard errors are reported in parentheses, which are clustered by zone in column (1), (4), (7) and by small zone in column (2), (3), (5), (6), (8) and (9). * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Table 6 Local Manufacturing Employment in a Vicinity of Industrial Parks

Dependent variable: Density of manufacturing employment in zone k,. *ZIP regressions*

	All zones		
	ZIP	ZIP with IVs	ZIP
	(1)	(2)	(3)
log($DISTANCE\_CENTER_k$)	-0.822*** (0.057)	-0.447*** (0.066)	-0.441*** (0.138)
$PARK_{kj}$	0.454 (0.400)	0.016 (0.123)	0.417 (0.414)
$AFTER_{jt}$	0.028 (0.157)	0.096** (0.046)	-0.072 (0.136)
$PARK_{kj}$*$AFTER_{jt}$	0.783** (0.367)	1.133*** (0.167)	0.180 (0.912)
log($DISTANCE\_CENTER_k$)* $PARK_{kj}$*$AFTER_{jt}$			-0.138 (0.226)
$PARK\_SIZE_j$* $PARK_{kj}$*$AFTER_{jt}$			0.017*** (0.006)
SOE_SHARE_{jt}* $PARK_{kj}$*$AFTER_{jt}$			-1.743*** (0.615)
$HUMAN\_CAPITAL_{jt}$* $PARK_{kj}$*$AFTER_{jt}$			4.256** (1.654)
$COAGGLOMERATION_{jt}$* $PARK_{kj}$*$AFTER_{jt}$			2.644 (2.208)
Constant	7.404*** (0.311)	5.028*** (1.538)	4.998*** (0.446)
District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Joint F-test of IVs in first stage (p-value)			
$PARK_{jt}$ regression		4.91*** (0.000)	
$PARK_{ij}$·$AFTER_{jt}$ regression		5.39*** (0.000)	

	Zones outside parks		
	ZIP	ZIP with IVs	ZIP
	(4)	(5)	(6)
log($DISTANCE\_CENTER_k$)	-0.394** (0.160)	-0.595*** (0.067)	-0.436*** (0.145)
log($DISTANCE\_PARK_{kjt}$)	-0.517*** (0.128)	-0.369** (0.187)	0.358 (0.386)
log($DISTANCE\_CENTER_j$)*log($DISTANCE\_PARK_{kjt}$)			-0.117 (0.075)
$PARK\_SIZE_j$*log($DISTANCE\_PARK_{kjt}$)			-0.006** (0.003)
SOE_SHARE_{jt}*log($DISTANCE\_PARK_{kjt}$)			0.766** (0.324)
$HUMAN\_CAPITAL_{jt}$*log($DISTANCE\_PARK_{kjt}$)			-4.399** (1.772)
$COAGGLOMERATION\_INDEX_{jt}$*log($DISTANCE\_PARK_{kjt}$)			-0.050 (0.124)
$GLOBAL\_PARK\_IMPACT_{kt}$	4.01e-4 (0.003)	0.003 (0.004)	-4.34e-4 (0.003)
Constant	5.505*** (0.645)	6.832*** (0.514)	6.101*** (0.730)
District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Joint F-test of IVs in first stage		9.29*** (0.000)	

Joint F-test of interaction variables			23.48^{***}				24.60^{***}
(p-value)			(0.000)				(0.000)
Observations	15999	13685	15999	Observations	15089	12958	15089
Nonzero obs.	4944	3976	4944	Nonzero obs.	4887	3965	4887
Vuong	26.09	-	25.81	Vuong	23.08	-	23.41

Note: The dependent variable is manufacturing jobs per square kilometer by zone by year. Column (1)-(3) report results from fitting versions of equation (16), which is a DID specification. The construction of IVs for *PARK* in column (2) is similar with that in Table 2. Column (4)-(6) report results from fitting versions of equation (17), and we drop those zones located in industrial parks. The construction of IVs for log(*DISTANCE_PARK*) (the distance from the zone's centroid to the closest park) in column (5) is similar with that in Table 3. In the first stage of the ZIP models in column (2) and (5), inflate regression, we regress employments density (in zone or sub-zone) on some location variables (*DISTANCE_CENTER*, $DISTANCE\_CENTER^2$, $DISTANCE\_CENTER^3$, log(*DISTANCE_RAILWAY*), log(*DISTANCE_AIRPORT*), log(*DISTANCE_UNIVERSITY*)), zone (or sub-zone) size, the city quadrant this zone (or sub-zone) locates in (north, south, east, or, west), district fixed effects and year fixed effects. For ZIP IV regressions (Column (2) and (5)) we employ bootstrap procedure to correct the standard errors). The Vuong statistics all favor the ZIP model. Observations in Shenzhen and Xi'an are dropped in column (2) and (5). Standard errors are reported in parentheses, which are clustered by district. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Table 7 Local Real Estate Construction and Pricing

	New home sales (by Grid) ZIP regressions				New home prices OLS regressions		
	ZIP	ZIP with IVs	ZIP		OLS	IV	OLS
	(1)	(2)	(3)		(4)	(5)	(6)
$\log(DISTANCE\_CENTER_w)$	-0.028	-0.020	-0.013	$\log(DISTANCE\_CENTER_l)$	-0.163^{***}	-0.170^{***}	-0.191^{***}
	(0.070)	(0.041)	(0.040)		(0.017)	(0.020)	(0.028)
$\log(DISTANCE\_PARK_{gjt})$	-0.221^{***}	-0.199^{***}	-0.075	**$\log(DISTANCE\_PARK_{ljt})$**	-0.101^{***}	-0.095^{***}	-0.161^{***}
	(0.050)	(0.051)	(0.107)		(0.012)	(0.013)	(0.043)
$\log(DISTANCE\_CENTER_j)*\log(DISTANCE\_PARK_{gjt})$			-0.054^{*}	$\log(DISTANCE\_CENTER_j)*\log(DISTANCE\_PARK_{ljt})$			0.015
			(0.030)				(0.013)
$PARK\_SIZE_j*\log(DISTANCE\_PARK_{gjt})$			-0.007^{**}	$PARK\_SIZE_j*\log(DISTANCE\_PARK_{ljt})$			-0.014^{*}
			(0.003)				(0.008)
$SOE\_SHARE_{jt}*\log(DISTANCE\_PARK_{gjt})$			-0.037	$SOE\_SHARE_{jt}*\log(DISTANCE\_PARK_{ljt})$			0.058^{**}
			(0.077)				(0.028)
$HUMAN\_CAPITAL_{jt}*\log(DISTANCE\_PARK_{gjt})$			-0.073^{**}	$HUMAN\_CAPITAL_{jt}*\log(DISTANCE\_PARK_{ljt})$			-0.084
			(0.032)				(0.079)
$COAGGLOMERATION_{jt}*\log(DISTANCE\_PARK_{gjt})$			0.098	$COAGGLOMERATION_{jt}*\log(DISTANCE\_PARK_{ljt})$			-0.159^{***}
			(0.190)				(0.058)
$GLOBAL\_PARK\_IMPACT_l$	0.023^{***}	0.024^{***}	0.024^{***}	$GLOBAL\_PARK\_IMPACT_l$	0.005^{***}	0.005^{***}	0.005^{***}
	(0.008)	(0.008)	(0.005)		(0.001)	(0.001)	(0.001)
				FAR_l	0.012^{***}	0.016^{***}	0.012^{***}
					(0.002)	(0.003)	(0.002)
				$GREEN_l$	$3.55e\text{-}3^{***}$	$3.87e\text{-}3^{***}$	$3.53e\text{-}3^{***}$
					(4.63e-4)	(5.20e-4)	(4.65e-4)
				$PARKING_l$	0.118^{***}	0.111^{***}	0.118^{***}
					(0.017)	(0.017)	(0.017)
Constant	2.861^{***}	-3.116	2.743^{***}	Constant	8.014^{***}	7.440^{***}	8.927^{***}
	(0.962)	(2.093)	(0.666)		(0.163)	(0.104)	(0.181)
District fixed effects	Yes	Yes	Yes	District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Year fixed effects	Yes	Yes	Yes
Joint F-test of IVs in first stage		3.51^{**}		Joint F-test of IVs in first stage		6.37^{***}	
(p-value)		(0.015)		(p-value)		(0.000)	
Joint F-test of interaction terms			11.90^{**}	Joint F-test of interaction terms			2.46^{**}
(p-value)			(0.036)	(p-value)			(0.032)
Observations	174280	153048	174280	Observations	117132	104367	117132

Nonzero obs.	164210	144113	164210	Nonzero obs.			
Vuong	85.36	-	85.36	Vuong			
R^2				R^2	0.721	0.710	0.722

Note: Column (1)-(3) report results from fitting versions of equation (18). The dependent variable is the count of new home sales by 2×2km grid by year. The construction of log(*DISTANCE_PARK*) (the distance from the grid's centroid to the closest park) in column (2) is similar with that in Table 3. Standard errors are reported in parentheses, which are clustered by zone. Column (4)-(6) report results from fitting versions of equation (19). The hedonic model is a regression of each home sale's price on its location attributes including distance to the nearest industrial park, distance to the city center, and physical attributes including floor area ratio, greening space ratio and parking space share (see more details in table 1). The construction of log(*DISTANCE_PARK*) (the distance from the housing complex to the closest park) in column (5) is similar with that in Table 3. For ZIP IV regressions (Column (2)) we employ bootstrap procedure to correct the standard errors). Robust standard errors are reported in parentheses, which are clustered by grid (2×2km). * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Table 8 Local Retail Establishment Growth

ZIP model, by grid by year

	RESTAURANT$_{kt}$			*ENTERTAINMENT*$_{kt}$			*SHOP*$_{kt}$		
	ZIP	ZIP with IVs	ZIP	ZIP	ZIP with IVs	ZIP	ZIP	ZIP with IVs	ZIP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(*DISTANCE_CENTER*$_w$)	-1.096***	-1.082***	-1.088***	-0.955***	-0.971***	-0.965***	-1.154***	-1.118***	-1.158***
	(0.090)	(0.086)	(0.091)	(0.069)	(0.051)	(0.063)	(0.093)	(0.048)	(0.004)
log(*DISTANCE_PARK*$_{gjt}$)	-0.349***	-0.292***	0.033	-0.274***	-0.254***	0.042	-0.346***	-0.264***	0.058***
	(0.040)	(0.039)	(0.223)	(0.036)	(0.054)	(0.180)	(0.046)	(0.046)	(0.017)
log(*DISTANCE_CENTER*$_j$)*log(*DISTANCE_PARK*$_{gjt}$)			-0.070*			-0.028			-0.0417***
			(0.041)			(0.033)			(0.003)
PARK_SIZE$_j$*log(*DISTANCE_PARK*$_{gjt}$)			-0.038*			-0.051***			-0.052***
			(0.020)			(0.017)			(0.002)
SOE_SHARE$_{jt}$*log(*DISTANCE_PARK*$_{gjt}$)			0.018			0.023***			0.009***
			(0.014)			(0.007)			(0.001)
HUMAN_CAPITAL$_{jt}$*log(*DISTANCE_PARK*$_{gjt}$)			-0.013			-0.016*			-0.015***
			(0.010)			(0.009)			(7.76e-4)
COAGGLOMERATION$_{jt}$ *log(*DISTANCE_PARK*$_{gjt}$)			-3.86e-4**			-2.28e-4*			-4.30e-4***
			(1.94e-4)			(1.33e-4)			(1.85e-5)
GLOBAL_PARK_IMPACT$_g$	0.006*	0.012**	0.004	-8.67e-4	0.005***	-0.002	0.005	0.008**	0.003***
	(0.004)	(0.006)	(0.003)	(0.003)	(0.001)	(0.003)	(0.004)	(0.004)	(2.90e-4)
Constant	6.145***	0.603	6.481***	4.980***	1.776***	5.339***	6.713***	2.703***	7.069***
	(0.612)	(1.532)	(0.625)	(0.521)	(0.354)	(0.519)	(0.703)	(0.159)	(0.049)
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Joint F-test of IVs in first stage		9.60***			15.16***			5.11***	
(p-value)		(0.000)			(0.000)			(0.000)	
Joint F-test of interaction terms			12.67**			21.11***			10.24*
(p-value)			(0.027)			(0.001)			(0.069)
Observations	174280	153048	174280	174280	153048	174280	174280	153048	174280
Nonzero obs.	156940	137888	156940	162129	142333	162129	157038	137698	157038
Vuong	41.95	-	36.14	28.36	-	24.45	40.34	-	37.29

Note: This table reports results from fitting versions of equation (20). The dependent variables are the count of restaurants, entertainment facilities and retail shops by grid by year, respectively. The construction of log(*DISTANCE_PARK*) (the distance from the grid's centroid to the closest park) in column (2) is similar with that in Table 3. For ZIP IV regressions (Column (2), (5) and (8)) we employ bootstrap procedure to correct the standard errors). Standard errors are reported in parentheses, which are clustered by zone. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Appendix A
Constructing Plant-level TFP and Park-level Co-agglomeration Index Measures

Constructing plant-level TFP measure

To estimate the plant-level TFP, we first consider a firm with a Cobb-Douglas production function,

$$Y_{it} = A_{it} L_{it}^{\alpha_l} K_{it}^{\alpha_k} \qquad \text{(A1)}$$

Where Y_{it} is output for plant *i* at time *t*, which is the function of labor, *L*, and capital, *K*. $A(\tau)$ measures the extent to which local public infrastructure, τ, enhances a plant's productivity.

Then we take the natural logs of equation (A1) and estimate:

$$\ln Y_{it} = \alpha_0 + \alpha_l \cdot \ln L_{it} + \alpha_k \cdot \ln K_{it} + \varepsilon_{it} \qquad \text{(A2)}$$

The dependent variable is real value added at the plant level, which is constructed by separately deflating output, net of goods purchased for resale and indirect taxes, and material inputs. Labor input, *L*, is measured by the number of employees each year. Capital stock is measured by plant-level real book value of fixed assets. The information on plants' fixed investment is not available in the ASIFs dataset, which reports the value of plants' fixed capital stocks at original purchase price, and their capital stock at original purchase prices less accumulated depreciation. These book values are the sum of nominal values and are not comparable across time and firms. We follow the procedure developed by Brandt et al (2012) and estimate plants' real value of the capital stock in each year. ε_{it} is the error term, which has two components, a white noise component, η_{it}, and a time-varying productivity shock, ω_{it}.

A major concern in estimating equation (A2) is that the correlation between unobservable productivity shocks, ω_{it}, and the input factors chosen by the plant. This may produce inconsistent estimates under OLS. A second problem is the endogeneity arising from sample selection. Plants exit when productivity falls below a certain threshold, and thus the surviving firms will have ω_{it} from the selected sample, which has an effect on the inputs employed. To address both of these two problems, we rely on Olley-Pakes (OP) estimator (Olley and Pakes 1996). We estimate the production function for plants in each two-digit sectors separately with Olley-Pakes. Thus, the measured TFP of plant *i* in year *t*, tfp_{it}, is defined as:

$$\log(TFP_{it}) = \ln Y_{it} - \hat{\alpha}_l \cdot \ln L_{it} - \hat{\alpha}_k \cdot \ln K_{it} \qquad \text{(A3)}$$

The estimated inputs coefficients obtained from estimating equation (A2) with Olley-Pakes are presented in Table A1. For comparison we also report OLS estimates with different production coefficients. The correlation between the OP and OLS is quite high: 0.976 in value. Both of the labor and capital estimates are underestimated with OLS, as expected when unobserved productivity is also possibly correlated with labor inputs and real capital stock.

Table A1 Firm-level production estimation results

	OP	OLS
Labor (L)	0.375***	0.351***
	(36.48)	(186.55)
Capital (K)	0.524***	0.417***
	(8.69)	(141.18)
Observations	249,267	247,378

Note: t statistics are reported in parentheses. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

Constructing park-level EG co-agglomeration index

We construct the EG co-agglomeration index in the following three steps: First, we quantify industry pairwise co-agglomeration by making use of the index of co-agglomeration (EG) developed in Ellision and Glaeser (1997). The EG index for industries *r* and *q* is:

$$\gamma_{rq} = \frac{\sum_{n=1}^{N}(s_{nr} - x_n)(s_{nq} - x_n)}{1 - \sum_{n=1}^{N} x_n^2} \qquad \text{(A4)}$$

Where n index geographic areas. s_{nr} is the share of industry *r*'s employment in area *n*. x_n is the mean employment share in area *n* across manufacturing industries.

Second, we calculate the employment share of industry *r* contained in the industrial park *j*, defined as s_{rj}.

Third, we compute industry pair employment shares in park *j* as follows:

$$s_{rq}^{j} = (s_{jr} + s_{jq})/2 \qquad \text{(A5)}$$

Finally, the co-agglomeration index for park *j* is

$$\gamma^{j} = \sum_{r} \sum_{q} \gamma_{rq} \cdot s_{rq}^{j} \qquad \text{(A6)}$$

In this analysis, we calculate EG index γ_{rq} for three-digit level manufacturing industries for each park using the ASIFs data set for the period of 1998-2007.

Appendix B Constructing Park-Vicinity Synergy Indices

We aim to identify the sources of an industrial park's spillover effect on the nearby incumbent firms. Following Glaeser and Kerr (2009), Ellison et al. (2010), and Jofre-Monseny et al.(2011), our first step is to construct measures of the extent to which two industries (1) have a strong input-and-output linkage; (2) share the same type of workers ; or (3) have a higher probability of knowledge spillovers. The second step is, for a specific industry an outside-park firm belongs to, we construct industry pairs by matching this industry with all the industries within the park, and then computing the weighted sum of the above industry-pair measures (the weight is set to be the employment share of that within-park industry in the whole park's employment).

We first measure the advantage that a firm exploits from locating close to a park rife with potential input suppliers and output demanders. The variables $Input_{r \leftarrow q}$ is defined as the share of industry *r's* inputs come from industry *q*, and $Onput_{r \rightarrow q}$ as the share of industry *r's* goes to industry *q*. These two measures range from zero to one since they are calculated relative to all input suppliers and output demanders. We calculate these two variables for two-digit-level manufacturing sectors using the input-output flows information according to the 2002 China Input-Output Tables published by China Bureau of Statistics.

Based on these two variables, the first synergy index measures the impact of a park *j* in terms of input flows on an industry *r* outside the park, which is defined as:

$$INPUT\_LINKAGE_{rj} = \sum_{q=1...Q} (Input_{r \leftarrow q} \cdot \frac{E_{qj}}{E_j}) \qquad (A7)$$

Similarly, the second synergy index measures the impact of a park *j* associated with output flows on an industry *r* outside the park, which is defined as:

$$OUTPUT\_LINKAGE_{rj} = \sum_{q=1...Q} (Onput_{r \rightarrow q} \cdot \frac{E_{qj}}{E_j}) \qquad (A8)$$

Where *Q* indexes the number of industries and *E* indexes the number of workers.

Now we turn to the labor pooling mechanism. According to the 1995 Industry Census data, employees are classified in 5 occupations, workers and apprentices, engineers, administrative staffs, logistics staffs, and others. Following Jofre-Monseny et al.(2011), we construct the labor similarity index for industry *q* and *r*, which measures the similarity of their distribution of workers by occupation between these two industries:

$$Labor\ Similarity_{rq} = 1 / \frac{1}{2}\sum_{O} \left| \frac{E_{or}}{E_r} - \frac{E_{oq}}{E_q} \right| \qquad \text{(A9)}$$

where O indexes the types of occupations and E indexes the number of workers. *Labor Similarity*$_{rq}$ represents the extent to which the share of workers in industry r need to change to mimic the distribution of occupations in industry q, ranging between 0 and 1. Hence, *Labor Similarity*$_{rq}$ takes the positive values and is greater than one. Based on labor similarity index, we calculate industry-specific the labor similarity weights for industry r with industries q as:

$$W_{rq} = \frac{Labor\ Similarity_{rq}}{\sum_{q=1...Q} Labor\ Similarity_{rq}} \qquad \text{(A10)}$$

Using W_{rq}, we define the availability of suitable labor forces of a park j for an industry r outside the park as our third synergy index:

$$LABOR\_POOLING_{rj} = \sum_{q=1...Q} (W_{rq} \cdot \frac{E_{qj}}{E_j}) \qquad \text{(A11)}$$

LABOR_POOLING$_{rj}$ is the weighted sum of industry r in park j employment shares where industries share the more similar types of labors in terms of occupations are given higher weights.

The last metric is to examine the extent to which a firm can enjoy knowledge spillover benefits from a park nearby. We use the 1995 Industry Census data and calculate the share of employees in terms of their education levels for 2-digit-level industries. We classify the workers into four types according to their educations, those with college and above, with high school, with middle school, and with primary school and below. The variable *Skill Similarity*$_{qr}$ measures the extent to which industry r and industry q share the same skilled workers:

$$Skill\ Similarity_{rq} = 1 / \frac{1}{2}\sum_{n} \left| \frac{E_{nr}}{E_r} - \frac{E_{nq}}{E_q} \right| \qquad \text{(A12)}$$

where n indexes education levels and $\frac{E_{nr}}{E_r}$ denotes the share of workers with education level n for industry r. *Skill Similarity*$_{rq}$ takes the positive values great than one. The industry-specific the labor similarity weights for industry r with industries q as:

$$S_{rq} = \frac{Skill\ Similarity_{rq}}{\sum_{q=1...Q} Skill\ Similarity_{rq}} \quad \text{(A13)}$$

Using S_{rq}, the knowledge spillovers in terms of skill sharing of a park j for an industry r outside the park is defined as our forth synergy index:

$$KNOWLEDGE\_SPILLOVER_{rj} = \sum_{q=1...Q} (S_{rq} \cdot \frac{E_{qj}}{E_j}) \quad \text{(A14)}$$

$KNOWLEDGE\_SPILLOVER_{rj}$ is the weighted sum of industry r in park j employment shares where industries share the more similar skilled workers are given higher weights.

Appendix C Probit Model of Plant Closing Probability

Table A2 Probit Model of Plant Closing Probability

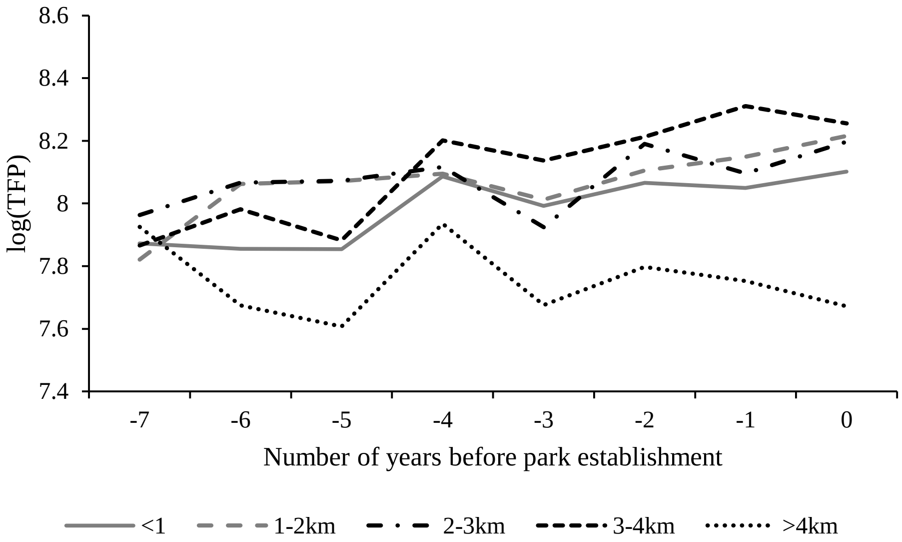
Dependent variable: Whether plant *i* is closed

	All	Lower TFP plants	Higher TFP plants
	(1)	(2)	(3)
AGE$_j$	0.004***	0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
log(*DISTANCE_PARK*$_i$)	**0.062****	**0.038**	**0.064***
	(0.029)	**(0.026)**	**(0.036)**
log(*DISTANCE_CENTER*$_i$)	-0.073**	-0.081**	-0.089***
	(0.032)	(0.035)	(0.034)
log(*DISTANCE_RAILWAY*$_i$)	-0.018	-0.025	-0.020
	(0.028)	(0.027)	(0.035)
log(*DISTANCE_AIRPORT*$_i$)	0.056	0.031	0.083*
	(0.036)	(0.039)	(0.043)
log(*DISTANCE_UNIVERSITY*$_i$)	-0.148***	-0.151***	-0.167***
	(0.028)	(0.029)	(0.034)
Constant	0.545***	0.949***	0.269
	(0.132)	(0.143)	(0.169)
City fixed effects	Yes	Yes	Yes
Observations	37527	18761	18766
R^2	201.9	177.8	95.90

Note: This table reports results from fitting versions of equation (13). The observation is a plant outside industrial parks. If the plant was closed during our study period (1998-2007), the dependent variable equals to 1, 0 otherwise. The distance variables are calculated as the average distances of plant *i* to those locations during its duration period. Liner probability regression is employed. Marginal effects are reported for *PROBIT* models (column (1) and (2)). Robust standard errors are reported in parentheses, which are clustered by district. * denotes $p < 0.10$, ** denotes $p < 0.05$, *** denotes $p < 0.01$.

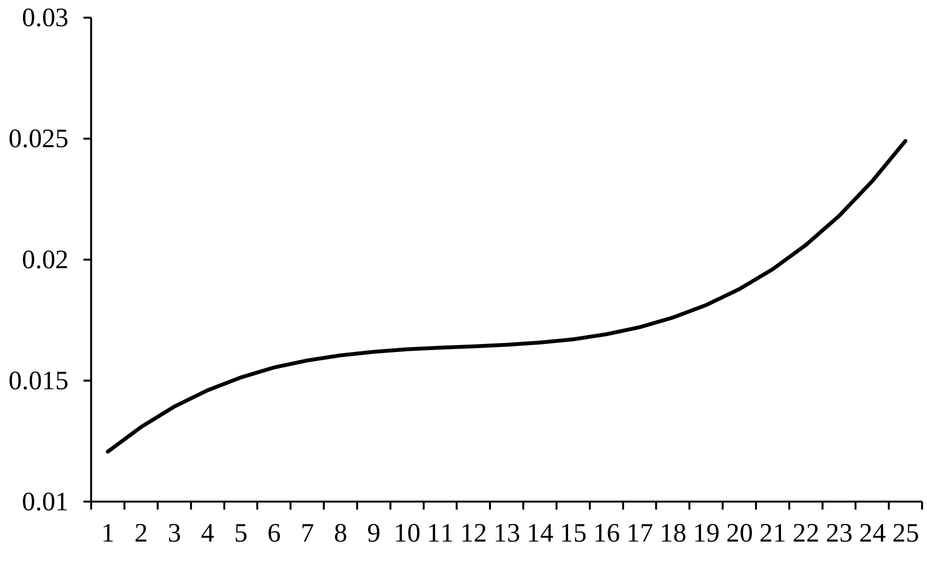
Appendix D. Figures

Figure D-1 TFP trends of incumbent firms before park establishment



Notes: We only include those incumbent firms which had existed at least two years before the introduction of an industrial park nearby. We do not observe any TFP drop before park establishment.

Figure D-2. Co-agglomeration Trend with the Age of an Industrial Park



$$\underset{}{Co-Agglomeration\ index} = \underset{(1.86^{*})}{3.19\times10^{-6}\cdot Age^{3}} - \underset{(-1.61)}{1.11\times10^{-4}\cdot Age^{2}} + \underset{(1.54)}{0.00133\cdot Age} + park\ \ fixed\ effects$$